OM nucleic - nucleic search, using sw model

Run on: November 9, 2003, 08:46:45 ; Search time 4837 Seconds
                                           (without alignments)
                                           10605.886 Million cell updates/sec

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Title:
Perfect score: 1254
Sequence: 1 atggctcttgctaggaagtcga.........cattgcggaagtgtggaag 1254
Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :

GenEmbl:*
  1: gb_ba:*
  2: gb_htg:*
  3: gb_in:*
  4: gb_om:*
  5: gb_ov:*
  6: gb_pat:*
  7: gb_ph:*
  8: gb_pl:*
  9: gb_pr:*
 10: gb_ro:*
 11: gb_sts:*
 12: gb_sy:*
 13: gb_un:*
 14: gb_vi:*
 15: gb_ba:*
 16: em_fun:*
 17: em_hum:*
 18: em_in:*
 19: em_mu:*
 20: em_om:*
 21: em_or:*
 22: em_ov:*
 23: em_pat:*
 24: em_ph:*
 25: em_pl:*
 26: em_ro:*
 27: em_sts:*
 28: em_un:*
 29: em_vi:*
 30: em_vrt:*
 31: em_htg_hum:*
 32: em_htg_inv:*
 33: em_htg_mus:*
 34: em_htg_pln:*
 35: em_htg_rod:*
 36: em_htg_mam:*
 37: em_htg_vrt:*
 38: em_sy:*
 39: em_htgo_hum:*
 40: em_htgo_mus:*
 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	855.4	68.2	1639	8	FSY251819	Fagus syl
2	826	65.9	1442	8	AF123395	Arabidops
3	826	65.9	1472	8	AY065174	Arabidops
4	826	65.9	1547	8	AB044348	Arabidops
5	825.8	65.9	1540	8	AB035272	Matricari
6	811.4	64.7	1516	8	AB052887	Oryza sat
7	811.4	64.7	1625	8	AB033537	Oryza sat
8	810.6	64.6	1275	6	AX653352	Sequence
9	810	64.6	1291	8	AY079403	Arabidops
10	810	64.6	1475	8	AY087503	Arabidops
11	810	64.6	1512	8	AY039965	Arabidops
12	779	62.1	1747	8	AF220199	Pinus tae
13	775.6	61.9	1553	8	AY011124	Dactylis
14	647.8	51.7	1287	10	RAT29	Rat liver m
15	647.8	51.7	1288	10	D83521	Rat liver m
16	643.8	51.3	1255	3	AB000491	Rattus no
17	631.8	50.4	1326	9	U43728	Manduca sex
18	631.8	50.4	1336	9	BC001932	Homo sapi
19	630.2	50.3	1277	10	BC002367	Homo sapi
20	628.6	50.1	1277	9	MMSUGIMR	M.musculus
21	625.4	49.9	1218	6	HUMPSP45	Homo sapien
22	625.4	49.9	1289	9	E11889	cDNA encodi
23	622.6	49.6	1321	4	HUMTRIP1M	Homo sapien
24	622.6	49.6	1956	8	NF041812	Naegleria f
25	620.6	49.5	1197	4	SSRNA26SP	SPO2280 Schizosacch
26	615.2	49.1	2586	9	AK095836	Homo sapi
27	609	48.6	1349	3	DMU97538	Drosophila
28	608.4	48.5	1360	3	AF043734	Drosophil
29	608.4	48.5	1363	3	AY051732	Drosophil
30	604.2	48.2	1170	3	DDITATBPB	Dictyostel
31	594.6	47.4	1665	5	XLXSUG1	X.laevis XS
32	593.8	47.3	1282	3	L16579	Drosophil
33	592.2	47.2	1956	8	SP002280	SPO2280 Schizosacch
34	592.2	47.2	35694	8	SPBC23G7	S.pombe c
35	576.8	46.0	1434	2	AY119229	Drosophil
36	563.8	45.0	27140	3	AC013978	Drosophil
37	563.8	45.0	173508	3	AC011251	Drosophil
38	563.8	45.0	315335	3	AE003568	Drosophil
39	551.2	44.0	251762	3	AE014851	Plasmodiu
40	551.2	44.0	310779	2	AC005140	Plasmodiu
41	549.6	43.8	1786	6	PFBPM	Plasmodiu
42	547.8	43.7	1218	6	AX595354	Sequence
43	547.8	43.7	1800	8	YSCTBY1A	S.cerevisia
44	546.2	43.6	1671	8	SCYGL048C	S.cerevisia
45	542.2	43.2	1415	10	BC030840	Mus muscu

ALIGNMENTS

```
TITLE        Up-regulation by GA3 of a new member of the AAA family (FsA1), in
             dormant beechnuts (Fagus sylvatica L.) showing Mg2+-dependent
             ATPase activity
JOURNAL      Unpublished
REFERENCE    2  (bases 1 to 1639)
AUTHORS      Lorenzo,O._
TITLE        Direct Submission
JOURNAL      Submitted (14-DEC-1999) Lorenzo O., Fisiologia Vegetal, Universidad
             de Salamanca, Plaza de los Doctores de la Reina s/n, Salamanca
             37007, SPAIN
FEATURES             Location/Qualifiers
     source          1..1639
                     /organism="Fagus sylvatica"
                     /mol_type="mRNA"
                     /db_xref="taxon:28930"
                     /clone="FsA1"
                     /tissue_type="seeds"
                     /clone_lib="Fagsyl1"
                     /dev_stage="2 weeks ABA-treated seeds"
     gene            1..1639
                     /gene="a1"
                     83..1339
     CDS             /gene="a1"
                     /note="rat binding protein"
                     /codon_start=1
                     /product="26S proteasome subunit 8"
                     /protein_id="CAB63651.1"
                     /db_xref="GI:6599051"
                     /translation="MAIETKHPEAVPADETCSAKCSKQGEGLKQYYLQHIHEHQLQVR
                     QKTHNLNRLEAQRNEFNRSRVRMLREEQLLQEPGSYVGEVKVKVMGNKNVLIVKHSEGK
                     YVVDIDKNIDITKITTSTRVALRNDSYVLHLILPSKVDPLVNLMVEKVPDSTDYDMIG
                     GLDQQIREKIEVIELPIKHPELFLESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDC
                     TFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGGNGD
                     SEVQDRMLELINQLDGPFESIKGIKVLMRTNRIDILDQALLRPGRIDRKIEFPNPNRES
                     RDILKIHSRRMLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALTAERRVRLVTQ
                     EDFEMAVAKVMKKETEKNMSLRKLWK"
BASE COUNT       518 a      298 c      379 g      444 t
ORIGIN
```

```
Qy      Query Match       68.2%;      Score 855.4;    DB 8;    Length 1639;
        Best Local Similarity   81.6%;    Pred. No. 9.1e-231;
        Matches 1017;  Conservative   0;  Mismatches 221;   Indels   9;  Gaps   2;

Qy       17 TTGAACTGAAGCATCCGGCGGCCGGAGGCGGGCGTACCGGAGGCGAATTGCTCCGCCAAGCCCACCA   76
            || ||||| ||| ||||||||| ||||||||||||||||||| |||||||||||||||||| ||
Db       90 TCGAAACGAAACATCCGGAGCCGGTGCCCGGCGGACGAGCACTTCCTCGGCGAAGTGCTCCA  149

Qy       77 AGCACGGGGCCGGAGGGCCTCCCCCCACTACTATTCTCTCACACGAGCATCCAGCTCCTTC  136
            || ||||||||||| |||| ||||| |||| || ||| |||||||||||||||| ||||
Db      150 AGCAAGGGGGAGGGGTTCTCAAGCAGTACTATCTCCAAACACATTCAGCGACCACCAGCTCCAGG  209

Qy      137 TTCGCCAAAAGACTCATAACCCCTCAACCCGTCCTGAGGCTCCGAGAAAACGACCTCCAATTCTA  196
            |||||||||||| ||||| ||||||||||||||||||||| ||| |||| ||||||||||||
Db      210 TTCGCCAAAAGACCCCACCACTCCAACCCGCCGTTAACGCAGTTCAATTCTC  269

Qy      197 GGGGTGAGGAATGCTGCGCGGAGAAGAATTTACAGCTTCTGCGAGGACCCCGGCTTCTATGTCGGTG  256
            ||||||||||||| |||||||||||||||||||||||||||||| |||| ||||||| ||||
Db      270 GAGTGAGAATGCTAAGGGAGGAGACTACAGCTACTCCAGGATCATATGTTGGTG  329

Qy      257 AAGTTGTCAAAGTAATGGGCAAGGACAAAGTCCTTGTCAACGGTTCCACCGACGAAGGAAAAT  316
            ||| ||||||||| ||||||||||||| |||||||||||| |||||||||||||||| |
Db      330 AAGTTGTCAAAGTAATGGGGAAGAATAAGGTTTTAGTTGAAGGTTCAATCTGAAAGGGAAAAT  389

Qy      317 ATGTTGTTGACATTGACAAAAATATTGACATTACAAAGATTACTCCATCCACTACAGTTG  376
            |||||||| ||||||||||| || ||||||| |||||||||||||| ||| |||||||||
Db      390 ATGTTGTTGCATCGATAAAAATATTGACATCACAAAGATAACTCCATCAACACAAGAGTTG  449

Qy      377 CACTCCCGAACGACACAGTTAGTCTTCACTTAGTTCTCGCCAAGAATAAGTTGATCCATTGG  436
            ||| ||| |||||| ||||| || ||||||| |||||||| ||||||||| || |||||
Db      450 CTCTTCGCAAATGACAGCTTAGTGTGTCCTCATTTAATCTTGCCAAGCAAAGTTGATTCCTTGG  509

Qy      437 TCAATCTGATGAAAGTTGAGAGAAGTTCCGGATTCTACATATGACATATGATTGGTGGTTTAG  496
            || ||| |||||||| || |||||| ||||||||| ||||||||||| ||| |||||||| |
Db      510 TCAACCTAATGAAAGTTGAAAAGGTTCCTGATTCAACATATGACATGATTGGTGGGGTCTCTTG  569
```

```
Db      510 TCAACCTAATGAAAGTTGAAAAGGTTCCTGATTCAACATATGACATGATTGGGGGGTCTCTTG  569
            || ||||| ||||||||||||  |||||||||||| |||||||||||||||||| |||
Qy      497 ACCAGCAAATTAAAGAAAATAAAAGAGGTCATTGAGCTACCAATCAAAACATCCTGAGCTCT  556
            ||||||||||||||||||| ||||||||  |||||||||| ||||||| ||  |||
Db      570 ACCAGCAAATTAAAGACAATAAAGGAGGTCATTGACTTCCAATCACCCTGAATTAT  629

Db      570 ACCAGCAAATTAAAGACAATAAAGGAGGTCATTGACTTCCAATCACCCTGAATTAT  629
            ||||||||||| ||||||||||||||||||||||||||||||||||| |
Qy      557 TT---GAAAGTCTTGGAATTGCACAACCAAAAGGGTGTCCTGCTCCTATAGTCAGTA  613
            ||    ||||||||||||||||||||||||||||||  |||| || ||||||||
Db      630 TTCTGAGAGCTCTTGGAATAGCTCAACCAAAGGGTGTACTGCTAATATGGGCCACCTGGTA  689

Db      630 TTCTGAGAGCTCTTGGAATAGCTCAACCAAAGGGTGTACTGCTAATATGGGCCACCTGGTA  689
            ||||||||||| |||| ||||| |||||||||||||||||| ||| |||||||||||
Qy      614 CAGGGTAAAAACATTGTTGGGCTAGGGCAGGTGGCTCATCATACTGACTGTACATTCATCAGGG  671
            ||| ||||||| |||||||| |||||||| ||| ||| ||||| |||||| |||| |||||
Db      690 CAGGGGAAAAACGCGTTGTTGGCTAGGGCAGTGGCACATCATACACAGAATTGTACTTTCATCAGGG  749

Db      690 CAGGGGAAAAACGCGTTGTTGGCTAGGGCAGTGGCACATCATACACAGAATTGTACTTTCATCAGGG  749
            |||||||||||||| ||| |||||| ||||| ||| || ||||| |||  ||| || |||||
Qy      674 TGTCTGGTTCGAGTTAGTTCAGAAATACATTGGAGAAGTTTCTAGAATGGTCAGGGAAC  733
            |||  |||| ||||| |||| ||| ||||||||||||| ||| || |||||||| |||
Db      750 TTTCTGGTTCTTGAATTAGTTCAGAAGTAACATTGGAGAAGGTTCCAGAATGGTTAGGGAAC  809

Db      750 TTTCTGGTTCTTGAATTAGTTCAGAAGTAACATTGGAGAAGGTTCCAGAATGGTTAGGGAAC  809
            |||||||||| ||||||||||||||| ||| ||||||||||| ||||||||||| |||||
Qy      734 TTTTTTGTTATAGGCCAGGGGACACATGCTCCATCAATTATCTTCATCGGATGAAATTGACAGTA  793
            ||||| || || |||||||| ||||| || ||||||| ||||| || |||| |||  |||
Db      810 TTTTTCGTTATGTCCCAGGGACACATGCTCCACTCTATCATAATTATGGATGAAAATAGACAGTA  869

Db      810 TTTTTCGTTATGTCCCAGGGACACATGCTCCACTCTATCATAATTATGGATGAAAATAGACAGTA  869
            ||||||| ||| |||| ||||||||||| || |||||| |||||| |||||||||| ||||||
Qy      794 TTGGATCTGCTCCGGATGGAAATCTGGAAGTGGCACGGTGATTTGAAGAAGATTAGCAGCCTACTA  853
            ||||||||| | ||||| |||| |||||||| || || |||||||||||||| ||||| |
Db      870 TTGGATCTGTTGTCTGGATGAATCTCGTTAGTGGCAATGGTGATAGTAGGAAGTGCAGGAACCACA  929

Db      870 TTGGATCTGTTGTCTGGATGAATCTCGTTAGTGGCAATGGTGATAGTAGGAAGTGCAGGAACCACA  929
            ||||| |||||| ||||| |||||| ||||| |||||||| ||| ||||| ||| || |||||
Qy      854 TGCTGGAACTTCTCAACCAGTTGGATGGATTTGAAGCTTCAAAATAAGATCAAGGTTTTGA  913
            |||||||| |||||||||| |||||| ||| ||||||| ||| ||||| |||||| ||
Db      930 TGCTGGAGCTTCTCAACCAGCTGGATGGATTTGAAAGCATCAAAACAAATTAAGGTTCTTA  989

Db      930 TGCTGGAGCTTCTCAACCAGCTGGATGGATTTGAAAGCATCAAAACAAATTAAGGTTCTTA  989
            || |||||||||| ||||||||| || |||||||| ||||||  ||| ||| || |||
Qy      914 TGGCCACCAAACGGATTGATATCCTCGGATCGAAGCCCTCTTAGACCAGACGAGATAGACC  973
            || ||||| ||| |||||| || |||||| |||| ||| ||||||| || ||||||||
Db      990 TGCGTACAAATCGTATTGATAATTCCGATCAAGCTCCTCTTAGCGCACGAGGCGAATTGATA  1049

Db      990 TGCGTACAAATCGTATTGATAATTCCGATCAAGCTCCTCTTAGCGCACGAGGCGAATTGATA  1049
            ||  || ||| || ||||| || ||||| ||| ||||| |||| || |||| |||||
Qy      974 GGAAAATTGAATTTCCAACCCCTAATGAAGAGCTCTCGGCTGGAATATTTGAAAATCCATT  1033
            ||||| || ||||| ||||| ||||||||||| ||| ||||| ||||||||||| ||||
Db     1050 GGAAGATTGAATTTCCTAATCCTAACCGGGAGTCTCGATTTGATATTTTAAAAATACATT  1109

Db     1050 GGAAGATTGAATTTCCTAATCCTAACCGGGAGTCTCGATTTGATATTTTAAAAATACATT  1109
            |||| |||||||||| || ||||||||| |||| || ||||| ||||  |||| |||
Qy     1034 CTAGAAGAATGAAATTTAATGCGGTGGTGGCCATGGATTTGAAGAAGATTGCCGAGAAGAGGATGAATG  1093
            ||||||||| ||| |||| |||||||||  ||||| |||||||||| || ||||||||||||  ||
Db     1110 CAAGAAGAATGAATCTGATCGGTGGGATTGACCTCAAGAAAATAGCTGAGAAAATGAACG  1169

Db     1110 CAAGAAGAATGAATCTGATCGGTGGGATTGACCTCAAGAAAATAGCTGAGAAAATGAACG  1169
            || |||||| |||| || || |||| || ||| ||||||||||  |||| ||||||
Qy     1094 GAGCATCTGGTGCTGAACTTAAGGCTGTTTGCACTGAAGCTGGAATGTTTGCTTTGACTG  1153
            |  ||||||| |||||  ||||| ||| || ||||||||||||||||||| || |||||
Db     1170 GTGCATCTGGTGCAGCTCAAGGCGTGTTGCACAGAAGCTGGAATGTTTGCTCTGACTG  1229

Db     1170 GTGCATCTGGTGCAGCTCAAGGCGTGTTGCACAGAAGCTGGAATGTTTGCTCTGACTG  1229
            | |||||||| |||||  ||  || |||||||||||||||| |||||| ||||||
Qy     1154 AGCGGGAGG------GTACACCTGACCTGCAAGGCGGATTTTGAGATGGCCGTGGCGAAGGTGA  1207
            | |||| ||      ||  ||| || ||||| ||| |||  || || ||||| |||||
Db     1230 CTGAGAGGAGGGTTCGACTGGTAACACAGAAGAGATTTTGAGATGGCAGTAGCAAAGGTAA  1289

Db     1230 CTGAGAGGAGGGTTCGACTGGTAACACAGAAGAGATTTTGAGATGGCAGTAGCAAAGGTAA  1289
            |||| || ||| || ||||||| | ||||||| ||| ||||||||| || || |||||
Qy     1208 TGAAAAAGGAGACTGAAAAAAACATGTCAATGTCATTGCGTAAGCTTTGGAAGTTGTGGAAG  1254
            ||||| ||| ||||||| | || || ||| ||||| |||||||||| ||| ||||||
Db     1290 TGAAGGAGGAGAGCCGAGAAAAACATGTCATTGCGGAAGCTAATGGAAG  1336
```

```
RESULT 2
LOCUS       AF123395
DEFINITION  AF123395                1442 bp       mRNA       linear    PLN 02-JAN-2000
            Arabidopsis thaliana 26S proteasome AAA-ATPase subunit RPT6a
            (RPT6a) mRNA, complete cds.
ACCESSION   AF123395
VERSION     AF123395.1    GI:6652887
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1442)
  AUTHORS   Fu,H., Doelling,J.H., Rubin,D.M. and Vierstra,R.D.
  TITLE     Structure and function analysis of the six regulatory particle
            triple-A ATPase subunits RPT1-6 of the 26S proteasome particle from
            Arabidopsis thaliana
```

```
JOURNAL   Unpublished
REFERENCE 2  (bases 1 to 1442)
  AUTHORS   Fu,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-1999) Institute of Botany, Academia Sinica, 128,
            Sec. 2, Academy Rd., Nankang, Taipei 11529, Taiwan
FEATURES             Location/Qualifiers
     source          1..1442
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /clone="97B4"
     gene            1..1442
                     /gene="RPT6a"
     CDS             85..1302
                     /gene="RPT6a"
                     /codon_start=1
                     /product="26S proteasome AAA-ATPase subunit RPT6a"
                     /protein_id="AAF22526.1"
                     /db_xref="GI:6652888"
                     /translation="MEETCNVKGAAAKQGBGLKQYLQHIHELQRQLRQKTNNINRLE
                     AQRNLNSRVRMLREELQLLQEPGSVGEVKVMGKNKVLVKHPEGKYVVDIDKSID
                     ITKITPSTRVALRNDSVVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIK
                     EVIELPIHQPELFESLGITAQPKGVLLYGPPGTGKTLLARAVAHHTDCTPIRVSGSELV
                     QKYIGEGSRMVRELFVWAREHAPSIIFMDEIDSIGSARMESSGSGNGDSEVQRTMLELL
                     NQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPNPNEESRFDILKIHSRX
                     MNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRRERVRHVTQEDFEMAVAKVMKK
                     DYEKNMSLRKLWK"
BASE COUNT      437 a    256 c    367 g    382 t
ORIGIN
```

```
Query Match      65.9%;  Score 826;  DB 8;  Length 1442;
Best Local Similarity  81.0%;
Matches  961;  Conservative  0;  Mismatches  225;  Indels  0;  Gaps  0;
            Pred. No. 1.9e-222;

Qy    69 GCCCACCAAGCCAGGGCGAGGGCGCCTCCGCCACTACTATTCTCAACATCCACGAGCATCA 128
         | | ||||  |  ||||| | ||||| ||  ||  |  | |||||| |||| ||||| |||
Db   114 GGGCGGGAAACCAGGGGAGAGGGCTTAAGCCAGTACTATTCTCCAGCACCATCCATGAGCTCCA 173

Qy   129 GCTCCTCCTTCCGCCAAAAGACTCCATAACCTCACCGTCTCTGAGGCTCAGAGAAACGACCT 188
         || || |  ||| ||| || ||| ||||| ||||| |||||  |||| ||||   ||| ||
Db   174 GCCGCCAGCTTCGACAGAGACTAATAACCTCAATCCAGCCTTGAAGCTCAAAGGAATGAACT 233

Qy   189 CAATTCTAGGGTGGAGGAATGCCCGAGGAATTACAGTTTCTGCAGGAACCCGGCTCTTA 248
         ||||||| | ||| |  ||||| | ||| | |||| | ||| ||||| | ||| || |
Db   234 CAATTCTCGAGTGAGAAATGCTCAGAGAAGAGTTACAGCCTCCTTCAAGAACCTGGGTCCTA 293

Qy   249 TGTCGGGTGGAAGTTGTCAAAGTAATGGGCAAGAACAAAGCTCCTTGTCAGGTCCACCCAGA 308
         |||  || || ||| | ||| | ||||||||| || || || ||| || | || |  |||
Db   294 TGTGGGAGGAGAAGTTGGTAAAAGTTATGGGCAAAGACAACAAGGTCCTTGGTTCATCCAGA 353

Qy   309 AGGAAAATATTGTTGTTGACATTGACAAAAAATATTGACATTACTCCACCCAC 368
         |||  | || ||||| ||| ||||||||| || || |||||||   |  | ||
Db   354 GGGGAAGTATGTTGTTGATATTGACAAAAGTATAGACATTACCAAATCACTCCATCAAC 413

Qy   369 TAGAGTTGCACTCCGCAACGACAGTTAATGTTCTCACTTAGTTCTGCCAAGTAAAGTTGA 428
         |  ||||||| |  ||   |  | |  | ||||| || | |||| |||||||||| | ||
Db   414 GAGAGTTGCTCTTCGTAATGATAGCTATGTTCTTCCACCTGGTTCTCCAAGTAAAGTAGA 473

Qy   429 TCCATTGGTCAATCTGATGAAAAGTTGAGAAAGTTCCGATTCTACATATGACATGGATTGG 488
         ||| ||||| || ||  |||| |||||||| |||||  | ||  | |||||||| |||||
Db   474 TCCCTTGGTTAACCTTATGAAAGTTGAGAAGGTTCCAGACTCCACCTATGACATGATTGG 533

Qy   489 TGGTTTAGACAGCAAAATTAAAGAAAATTAAAGAGGTCATTGAGCTACCACCAATCAAACATTCC 548
         |||| | || ||||  || |||| || |||  |||||||||||||| || || ||| | |||||
Db   534 TGGTCTTGACCAGCAGATTAAGGAAATAAAGGAGGTCATTGAGCTGCCAATCACAAGCATCC 593

Qy   549 TGAGCTCGTTTGAAAGTCTTGGAATTGCACAACCAAAGGGTGTCCTGCTCTCAATGGGGCACC 608
         |||  | |||||| |||||| | |||||||| |||||||||||| || ||  | | |||| ||
Db   594 TGAATTGTTTGAGTCGTCTTGGCATTGCACCAAAGGGTGTGTTGTTATACGGTCCACC 653

Qy   609 TGGTACAGGTAAAACATTGTTGGCTAGGGCAGTGGCTCATCATACTGACTGTACATTCAT 668
         ||| || || || ||  | ||| |||||| ||||| |||||| ||||| ||  | ||||
Db   654 TGGAACTGGGAAGACAACTATGGCTCGGGCTGTTGCACATCACACTGACTTGCACTTCAT 713

Qy   669 CAGGGTGTCTGGTCTCTGAGTTAGTCAGAAATAGTCATTGAAGAGGTTCGAATGTCAG 728
         ||||||||| ||| | ||||||| | |||||| |||| | |||||||| | ||||| |
Db   714 AAGAGTTTCTCGGTTCCCGAGCTGTTGCCAGAAATACCATTGGAGAAGGTTCTAGAATGTCAG 773

Qy   729 GGAACTTTTTCGTTATGGCCAGGGAACATGCTCCATCAATATATTCTCAATGGATGAAATTGA 788
         ||||||||| | | ||||| ||||||| ||| | ||||| | ||| || ||| || |  | |
Db   774 AGAACTTTTTTGTGATGGCAAGGGAACCATGCAACATCAATCTTCAATTGGAATGGTCGCG 833

Qy   789 CAGTATTGGGATCTGCTCCGGATGGAATCTGGAAGTGGCCAACCGTGATAGTGAGGTACAGCG 848
         |||||||||  ||||| | |||||||||||||||||||   || |||| |||||||| || ||
Db   834 TAGTATTGGGTCTGCTCGGATGGAATCTGGAAGTGGAAGTGGACAGTGAGGTGCAACG 893

Qy   849 TACTATGCTCGGAACTTCTCAACACCAGTTGGATTGGAGGTTCAAATAAGATCTCAAGGT 908
         || ||||||| || ||||||| || |||| |  ||| ||  | ||| || |   ||||||
Db   894 GACCATGCTCGAGCTTCTCAATCAACTTGACGGTTTTGAAGCGTCAAACAAAATCAAGGT 953

Qy   909 TTTGATGGGCCACCAATCGGGATTGATATCTCGGGATTCAGCCCTCCTTAGACCAGGAGGAT 968
         |||||||| | ||  |||| |||||||||   |  | | ||||||| ||||| |||||  ||
Db   954 TTTGATGGCTACAAATCGGTATTGATATTGATCAAGCTCTTCTCAGGCCTGGAAGAAT 1013

Qy   969 AGACCGGAAAATTGAAATTCCAACCCAATGAAGAGAGTCTCGGCTGGATTATTTTGAAAAT 1028
         | |  ||||||||||| |||| | ||| |||||||||||||   | || | ||||||| ||
Db  1014 TGATAGGAAAATTGAATTTCCTAATCCTAATGAAGAGAGTCACGTTTTGACATCTTGAAGAT 1073

Qy  1029 CCATTCTAGAGAGATGAAATTAAATGAGCGTGCCATTGATTTGAAGAAGAATTGCCGAGAAGAT 1088
         |||| | || ||||||||||| |  |  |||| ||||| ||||||||||||||||| ||| ||
Db  1074 TCACTCGAGGAAATGAATTGATCGGTGAAATCGATCTGAAAAAAGATTGCCAGAGAGAAT 1133

Qy  1089 GAATGGAGCATCTGGTGCTGGACTTAAGGCTGTTGCACTGAAGCTGGAAATGGTTGCTTT 1148
         |||||| ||||||||||||||||||||||||||||||||||||||||||||||| |||| |
Db  1134 GAATGGCGCTTCAGGTGCTGGACTTAAGGCTGTTGCACTGAAGCTGGAAATGTTTGCTCT 1193

Qy  1149 GAGGGAGCGGGAGGGTACAGGTCGACTCAGGAGGATTTGAGATGGCGTGGCGGAAGGTGAT 1208
         | | ||||| ||| ||ACA ||||| ||  ||| ||||||||||||||| |  |||| |||
Db  1194 GCGTGAGAGGGAGAGTACACGTGAGCACAAGAAGAACTTTGAGATGGCGGCCAAGGTAAT 1253

Qy  1209 GAAAAAGGAGACTGAAAAAAAATGTCATTGCGGAAGTTGTGGAAG 1254
         ||| || || || |||||||| |||| |||| ||| |||| ||||
Db  1254 GAAGAAAGACACCGAGAAGAACATGTCTCTGCGTAAGCTGTGTGGAAG 1299
```

```
RESULT 3
AY065174

LOCUS       AY065174                1472 bp    mRNA    linear   PLN 10-DEC-2001
DEFINITION  Arabidopsis thaliana unknown protein mRNA, complete cds.
ACCESSION   AY065174
VERSION     AY065174.1  GI:17473853
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1472)
  AUTHORS   Nguyen,M., Karlin-Neumann,G., Southwick,A., Lam,B., Miranda,M.,
            Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H.,
            Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamiya,A.,
            Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K.,
            Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K.,
            Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-2001) Stanford University, 855 California Avenue,
            Palo Alto, CA 94304, USA
COMMENT     e-mail for correspondence: arab@sequence.stanford.edu

            RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
```

```
Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Nguyen,M.;
Southwick,A., Karlin-Neumann,G., Lam,B., Miranda,M., Palm,C.J.,
Bowser,L., Jones,T., Banh,J., Chen,H., Cheuk,R., Chung,M.K.,
Kim,C., Lin,J., Liu,S.X., Pham,P.K., Sakano,H., Shinn,P.,
Yamada,K., Ecker,J., Theologis,A. and Davis,R.W.

Nguyen,M. (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed
equally to this work. Shinozaki,K. (RIKEN GSC) and Davis,R.W.
(SSP/Stanford) contributed equally to this work as PIs.

FEATURES             Location/Qualifiers
     source          1..1472
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="V"
                     /clone="RAFL06-08-I01"
                     /note="This clone is in pBluescript
                     ecotype: Columbia"
     CDS             92..1351
                     /note="unannotated coding sequence from BAC F28116"
                     /codon_start=1
                     /product="unknown protein"
                     /db_xref="GI:17473854"
                     /protein_id="AAL38350.1"
                     /translation="MAAVGVDSRPETAMEETCNVKGAAAKQGEGLKQYYLQHHELQ
                     RQLRQKNNLRLEAQRNEINSRVRMLREELQLLQEPGSYVGEVVKVMGNKVLVKVH
                     PEGKYVVDIDKSIDIKIITSTRVALRNDSYLHLVLPSKVDPLVNLMKVEKVPDSTY
                     DMIGGLDQIKEIKEIVELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHH
                     TDCTFIRVSGSELVQKIIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGSG
                     NGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPNPN
                     EESRFDILKIHSRKMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHVT
                     QEDFEMAVAKVMKKDTEKNMSLRKLMK"
BASE COUNT      429 a    271 c    371 g    401 t
ORIGIN

Query Match             65.9%;  Score 826;  DB 8;  length 1472;  Pred. No. 1.9e-222;
Best Local Similarity   81.0%;
Matches  961;           Conservative    0;   Mismatches  225;   Indels    0;   Gaps    0;

Qy    69 GCCCACCAAGCCAGGGCGACGGGCCTCCGCGCCACTACTTACTTCTCTCTCAACATCCACGAGAGCATCA 128
            ||| |||||| |  |  |  |||  | ||||||||| ||||| ||| |||||||||||||| |||||||
Db   163 GGGCGGCGGAAACAAGGGGGAAGGGCTTAAGCAGTACTATCTCCAGCACATCCATAGAGCTCCA 222

Qy   129 GCTCCTTCTTCGCCAAAAGACTCATAACCTCAACCGTCTCGAGGCTCAGAGAAACGACT 188
            ||||| ||| |||   | |  | ||| || ||   ||| |||||||||||| ||||||
Db   223 GCGCCAGCTTCGACAGAAGGACTAATAACCTCAATCGCCCTTGAAGCTCAAAGTGAACT 282

Qy   189 CAAATTCTAGGGTGGAGGAATCCTGCCGGAGGAAGTTACCAGCTTCTCGCAGGAGAACCCGGCGTCTTA 248
            | |||||| ||||||  || || || |||  ||  || ||||| ||| ||||| ||| ||||| ||||
Db   283 CAATTCTCGAGTGAGAAGTCCTCAGAGAGAGTTACAGCTCCTTCCTTCAAGAACCTGGGTCCTA 342

Qy   249 TGTCGGTGGAAGTTGTCAAAGTAATGGCGAAGAACAAAAGTCCTTGCTCAAGGTCCACCCAGA 308
            || |||| ||| | ||| |  | |||||  |  |  |  | ||  | |||||||||  |||||
Db   343 TGTGGGAGAAGTGGTAAAAGTTATGGGAAGAGAACAAAGACAAGGTTCATTCATCCA 402

Qy   309 AGGAAAATATGTTGTTGACATTGACAAAAATATTGACATTACAAAGAATTACTCCATCCAC 368
            |||  || ||| || ||||  |||  |  | ||||||  || || || ||||||||||||
Db   403 GGGGAAGTATGTTGTTGATATTGACAAGAGTATTGACATTAAGATCATCACTTCCATCAAC 462

Qy   369 TAGAGTTGCACTCCGCACGACACGTTATGTTCTTCACTTAGTTCTGCCAAGTAAAGTTGA 428
            || || || |||  ||| |  | |||  | || ||||| ||| |||||||||||| |||
Db   463 GAGAGTTGCTCTTCGTAATGACAGTTACCTGCATCTTGTACTTCCTTCGAAAGTGGATCCA 522

Qy   429 TCCATTGGTCAATCTGATGAAAAGTTGAGAAAGTTCCCGGATTCTACATATGTGACATTGG 488
            || ||||| |||| |  ||  |  ||| | |   || || || |||   | || ||  ||
Db   523 TCCCCTGGTTAACCTTATGAAAGTTGAGAAGGTTCCAGACTCCACTTATGACATGATTGG 582

Qy   489 TGGTTTAGACCAGCAAAATTAAAGAAATAAAAGAGGTCATTGAGCTACCAATCAAACATCC 548
            ||||||||| ||| | |  |  | |  |   ||| ||||||| || || ||||| |  ||
Db   583 TGGTCTTGACCAGATTAAGGAAATAAAGGAGGTCATTGAGCTGCCAATCAAGCATCC 642

Qy   549 TGAGCTGTTGAAAGTCTTGGAATTGCACAACAAAAGGTGTCCTCGTGTAATGGTCAG 608
            ||| || |||||| |||||||||||||| |||| ||| || || |  ||| | || |||  |
Db   643 TGAATTGTTTGAGTCTCTGGAATTGGCCAGCATAGGGTGTGTGTTGTTATACGGTCCACC 702

Qy   609 TGGTACAGGTAAAACATTGTTGGCTAGGGCAGTGGCTCATCACTACTGACTGTTCAT 668
            |||||| | ||||  ||| ||||| |||||| || || | || | ||||| | |||
Db   703 TGGAACTGGGAAGACATTTCTGGGCAGGGCTGTGGCACACCACACTGACTGCACTTCA 762

Qy   669 CAGGGTGTCTGGTTCTGAGTTAGTTCAGAAAATACATTGGAAGAGGTTCTCAGAATGGT 728
            | ||| || | | | |||||| | |  || ||||| |||| ||| |||  ||||| |||
Db   763 AAGAGTTCCTGGTTCCGAGCTTGTCCAGAAAATACACGAGGGTTCCAGAATGGTCAG 822

Qy   729 GGAACTTTTTGGTTATGGGCCAGGGAACATCGCTCCATCAATTATCTTCATGGATGAAATTGA 788
            |||||| ||| | |||||||| ||||| |   |||| | |||||| ||| ||||| |  |||
Db   823 AGAACTTTTTGGTGATGGGCCAAGGGAACGCATCCATCATCTTCATGGATGAAATTGA 882

Qy   789 CAGTATTGGGATCTGCTCCGGATGGAATCTGAAGTGGCAACGGTGATGAGTAGTGCAGCG 848
            || || ||| || |  |||| |||||| |||| || || | || || ||  |||  |||
Db   883 TAGTATTGGGTCTGCTCGTATGGAATCTGGAAGTGGAAATGGTGACTCAGAAGTGCAGCG 942

Qy   849 TACTAYGCTGGAACTTCTCAACCAGTTGCATGGATTCAGAAGATTCAAAATAAGATCAAGGT 908
            |||  |||| ||||| |||  |||||| | ||| |||||||| || |||   |||||||||
Db   943 GACCATGCTGGAGCTTCTCAATCAACTTGACGGGTTTGAAGCGTCAAACAAAATCAAGGT 1002

Qy   909 TTTGATGGCCACCAATCGGATTGATATCCTGGATCAAGCCCTCCTTCGACCAGGACGGGAT 968
            ||| |||| ||||| || || | || || |||| ||||| ||| ||||| ||||| |||
Db  1003 TTTGATGGCTACAAATCGTATTGATATTCTGGATCAAGCGCTGCTTCGTCCAGGTCGGAT 1062

Qy   969 AGACCGGAAAATTGAAATTCCAAACCCCTAAATGAAAGAGTCTCGCGCTGGGATATTTTGAAAT 1028
            | ||||||| || ||| ||||  | |||| ||||||||| || || |  || || | |||||
Db  1063 TGATAGGAAAATTGAATTTCCTAATCCTAATGAAGAGAGTCACGTTTGACATCTTGAAGAT 1122

Qy  1029 CCATTCTAGRAAGTATGAAATTTAATGCGTGGCATCATGATTTGAATTGAAGAGAAGGAGAAGAGAT 1088
            ||  ||||||||||| || || ||||  ||| || ||||   || || | |||  | || ||
Db  1123 TCACTCGAGRAAAATGAATTTGATGCGTGGCATCGATCTGAAGAAGATTGCAGAGAAGAT 1182

Qy  1089 GAATGGAGCAATCTGGTGCTGACTTAAGGCGTGTTGCACTGAAGCTGGAAGTTGCTTT 1148
            |||  | |||||| |||| ||||  || |  |||||| || ||||||||  ||| ||
Db  1183 GAATGGAGCTTCAGGTGCTGAGCTCAAAGCTGTGTGCACGGAAGGTTGCTCT 1242

Qy  1149 GAGGGAGAGCGGGAGGTACACCGTACTGACGTCAGGAGGAGATTTGAGATGCCGGTGGGAAGGTGAT 1208
            ||||| |  ||||||| || |  | |||   ||||| ||||| |||| ||| ||  || ||||
Db  1243 GCGTGAGAGGAGAGTACCAGTGACACAGGAAGACTTTGAGATGGCGGTGGCCAAGGTAAT 1302

Qy  1209 GAAAAAGGAGAGCTGAAAAAAACATTCATTGCGAAGAGTTGTGGAAG 1254
            ||||| | |  ||| |  | |  | ||| || || | || || || ||
Db  1303 GAAGAAGACACGGAGAAGAACATGTCTCTGCGTAAGCTGATGAAG 1348
```

RESULT 4
AB044348

```
LOCUS       AB044348                1547 bp    mRNA    linear   PLN 04-APR-2001
DEFINITION  Arabidopsis thaliana AtSUG1 mRNA, complete cds.
ACCESSION   AB044348
VERSION     AB044348.1  GI:13537114
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (sites)
  AUTHORS   Kuromori,T. and Yamamoto,M.
  TITLE     Identification of a cDNA from Arabidopsis thaliana for a member of
            conserved Sug1 CAD family
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1547)
  AUTHORS   Kuromori,T. and Yamamoto,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2000) Takashi Kuromori, The Institute of Physical
            and Chemical Research (RIKEN), Plant Molecular Biology Laboratory;
            3-1-1 Koyadai, Tsukuba 305-0074, Japan
```

(E-mail:kuromori@rtc.riken.go.jp, Tel:81-298-36-4359,
Fax:81-298-36-6090)

```
FEATURES             Location/Qualifiers
source               1..1547
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
gene                 1..1547
                     /gene="AtSUG1"
CDS                  104..1363
                     /gene="AtSUG1"
                     /codon_start=1
                     /protein_id="BAB40755.1"
                     /db_xref="GI:13537115"
                     /translation="MAAVGVDSRRPETAMEETCNVKGAAAKQGEGLKQYLQHIHELQ
                     RQLRQKNNLNRLEAORNEINSRVRMLREELQLLQEEGGSYUGEVUVKMGNKNVLVKVH
                     PEGKYVVDIDKSIDIHIITPSTRVALRNDSVVLHLVLPSKVDPLVNLMKVEKVPDSTY
                     DMIGGLDQIKEIKEVIELPIKHPELFESLGIAQRKGVLLVGPPGTGKTLLARAVAHH
                     TDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSSARMESGSG
                     NGDSEVORTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPNPN
                     EESRFDLIKIHSRKMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHVT
                     QEDFEMAVAKVMKKDTEKNMSLRKLWK"
polyA_site           1521
                     /gene="AtSUG1"

BASE COUNT      468 a    278 c    384 g    417 t
ORIGIN
```

Query Match   65.9%;   Score 826;   DB 8;   Length 1547;
Best Local Similarity 81.0%;   Pred. No. 1.9e-222;
Matches 961; Conservative   0; Mismatches 225; Indels   0; Gaps   0;

```
Qy      69 GCCCACCAAGCACAGGGCGGCGAGGGCCCTCCGCGCCACTACTTACTTCTCGAACATCCACCGAGCATCA 128
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175 GGGGGCGGAAACAAGGGAGGAAGGGCTTAAGCAGTACTACATCTCCAGCACATCCAATGAGCTCCA 234

Qy     129 GCTCCTTCTTCGCAAAAGACTCATAACCTCAACCGTCTCGAGGCTCAGAGGAAACGACT 188
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     235 GCGCCAGCTTCGACAGAGAAGACTAANACCTCANTCGCCTGAAGCCTGAAAGGAATGAACT 294

Qy     189 CAATTCTAGGGTGAGGAATGCTGCCGGACGAAGAATTACAGGCTTCTGCCAGGAGAACCCGGCGTCTTA 248
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     295 CAATTCTCGAGTGAGAGAATGCTCAGAGAGAGTACAGCTCCTTCCTCAAGAACCTGGGTCCTA 354

Qy     249 TGTCCGGTGAAGTTGTCAAAGTAATGGGCCAAGAACAAAAGTCCTTGTTCAAGGTCCACCCAGA 308
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     355 TGTGGGAGAAGTGGTTAAAAGTTATGGGAAAGAACAAGGTCTTGGTTAAGGTTCATCCAGA 414

Qy     309 AGGAAAAATATGTTGTTGACATTGACATTACAAAGATTACTCCATCCAC 368
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     415 GGGGAAGTATGTTGTTGATATTGACAAAAGTATAGACAATTACGGAAAATCACTCCATCCAAC 474

Qy     369 TAGAGTTGCACTCCGACNCGAGCACNATTTATCTTCTTCACTTAGTTGTTCTGCCAAGTAAAGTTGA 428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     475 GAGAGTGCTTCTTCCAGTNTCTCCAAGTAAAGTNTCCCACCCTGGTTCTGCCAAGTAAAGTAGA 534

Qy     429 TCCAATTGGTCAATCTGAGAAAAGTTGAAGAAGTTTCCCGGATTCTACNATGACNATGATTGG 488
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     535 TCCCTTGGTTAACCTTATGAAAGTTGAGAAGGTTCCAGACTCCACCTATGACANTGATTGG 594

Qy     489 TGGGTTTAGACCAGCAAAATTAAAAGAAAATAAAAGNAGGTCAGATTGGCTACCAACTCANACATCC 548
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     595 TGGGTCTTGACCAGCAGATTAAGGAAANAATAAAAGGGAGGTCGATTGGCTGCCAANTCANAGCATCC 654

Qy     549 TGAGCTGTTTGAAAAGTTCTTGGAAATTGCACAACCAAAGGGGTGTCCTGCTCTANTAGGGCCACC 608
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     655 TGAATTGTTTGAGTCTCTTGGANATTGCCCAGCCAAAGGGTGNGTTGTTATTACGGTCCACC 714

Qy     609 TGGTACAGTAAAACATTGTTGGGTAGGGCCAGTGGCCTCATCANTACTGACTGTACATTCAT 668
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     715 TGGAACTGGAGAGACTATTGCTCGGGCTGTTGCACATCACTGACTGTACTTCAT 774

Qy     669 CAGGGTGTCTGGGTTCTGAGTTAGTTCAGAAAATACATTGGAGAAGGTTCTAGAATGGTCAG 728
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
RESULT 5
AB035272
LOCUS         AB035272             1540 bp    mRNA            linear   PLN 20-DEC-2002
DEFINITION    Matricaria chamomilla McTBP1 mRNA for TAT-binding protein homolog,
              complete cds.
ACCESSION     AB035272
VERSION       AB035272.2  GI:17385639
KEYWORDS      .
SOURCE        Matricaria chamomilla
  ORGANISM    Matricaria chamomilla
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
              Anthemideae; Matricaria.
REFERENCE     1
  AUTHORS     Ashida,Y., Nishimoto,M., Matsushima,A., Watanabe,J. and Hirata,T.
  TITLE       Molecular cloning and mRNA expression of geraniol-inducible genes
              in cultured shoot primordia of Matricaria chamomilla
  JOURNAL     Biosci. Biotechnol. Biochem. 66, 2511-2514 (2002)
REFERENCE     2  (bases 1 to 1540)
  AUTHORS     Ashida,Y. and Hirata,T.
  TITLE       Direct Submission
  JOURNAL     Submitted (26-NOV-1999) Yoshiyuki Ashida, Graduate School of
              Science, Hiroshima University, Department of Mathematical and Life
              Sciences; 1-3-1, Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526,
              Japan (E-mail:yashida@sci.hiroshima-u.ac.jp,
              Tel:81-824-22-7111(ex.2807), Fax:81-824-24-7435)
              On Dec 6, 2001 this sequence version replaced gi:6472926.
COMMENT
FEATURES             Location/Qualifiers
source               1..1540
                     /organism="Matricaria chamomilla"
                     /mol_type="mRNA"
                     /db_xref="taxon:98504"
gene                 1..1540
```

```
Db     775 AAGAGTTCTTGGTTCCGAGCTGTGTCCAGAAATACATTGGAGGAAGGTTCTAGAATGGTCAG 834

Qy     729 GGAACTTTTGTTTATGGGCCAGGGAACAATGGTCCCATTCCACCAATTATCTTCATGGAGAAAATTGA 788
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     835 AGAACTTTTGTGTGATGGGACAAGGGAGCATGCACCATCAATCANTCTTCATGGAGAAAATTGA 894

Qy     789 CAGTATTGGGATCTGCCTCGGATGGAATCTGGAAGNTGGCAACGGTGATANGTAGTGAGTGAGCTACAGCG 848
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     895 TAGTATTGGGTGCTCGCTCGTATGGAATCTGGAAGTGAAGTGAAATGGTGGACAGTGGCAGTGGTTACACG 954

Qy     849 TACTATGCTGGAACTTCTCAACCACGTTGGGATGGATTGGGAGGTTCAAAATAAGATCAAAGGT 908
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     955 GACCCATGCTCGAGCTTCTGAATCAACCAGTTGACGGGTTTGGAGGCGTCAAAACAAAAATCAAGGT 1014

Qy     909 TTTGATGGGCCACCAAATCGGGATTGAGATCTTCTGGGATCAAGGCCCTCCTTAGCACCAGGACGGGAT 968
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1015 TTTGATGGGCTACAAAATCGTATTGGGATCAAGGTCTGAAGCTCTTCCTCAGGGCGGAAGAAT 1074

Qy     969 AGACCCGGAAAATTGAATTCCAAACCCCTAATGAAGAAGAGTCTCCGGCTCTGGATATTTGAAAAT 1028
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1075 TGATAGGAAAATTGAATTCTCCTAATCCTAATGAACAGAGTCACGTTTTGAGATGGGGGTGGCCAAGGTAAT 1314

Qy    1029 CCATTCTAGAGAATGAAATTAATGCGTGGCCATTGATTGAAGAAGATTCCCGACAAGAT 1088
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1135 TCACTCCGAGGAAAATGAAATTTGATGGTCCGGTGAAATCGAATCGAAAAAAGAATTGCAGAGAAGAAGAT 1194

Qy    1149 GAGGGAGCGGAGGGGTGCTACACCTGACTCAGAGAGGATTTTGAGATGGCGCGGAAGGTGAT 1208
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1195 GAATGGCGCTTCAGGTGCTGCAGCTCGAAGGCTGTGTTCGCAGTGACAAGGAGGGAAGTGTTTCT 1254

Qy    1209 GAAAAAGGAGCTGAAAAAACATGTCATTGCGGAAGTTGTCGGAG 1254
           ||||||||||||||||||||||||||||||||||||||||||||
Db    1255 GCGTGAGAGAGGAGTCACAGTGACACAGAAGAGACTTTGAGATGGCGTGGCCGGAAGGTAAT 1314
Db    1315 GAAGAAGAAGACAAGGAAGAAGAACACATGTCTCTGCGTAAGCTGTGGAAG 1360
```

```
CDS            /gene="MtTBP1"
               121..1365
               /gene="MtTBP1"
               /codon_start=1
               /product="TAT-binding protein homolog"
               /protein_id="BAA87070.2"
               /db_xref="GI:17389640"
               /translation="MKPASSSADSGTESCSSKGVAKTCSEGLRQYYLQHIHDAQLQVRQ
               KTHNLNRLEAQRNELNSRVGMLKEELQLLQEPGSYVGEVKVRMGKSKVLVKRHPEGKY
               VVIDKSIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGG
               LDQQIKEVIELPIKKPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCAF
               IRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSE
               VQRTMLELLNQLDGFESSTKIKVLMATNRIDILDPALLRPGRIDREIEFPPNEDSRF
               DILKIHSRKMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFE
               MAVAKVMKKETEKNMSLRKLWK"

BASE COUNT    492 a    269 c    354 g    425 t
ORIGIN

Query Match   65.9%;   Score 825.8;  DB 8;  Length 1540;
Best Local Similarity   81.5%;   Pred. No. 2.2e-222;
Matches 956;  Conservative 0;  Mismatches 217;  Indels 0;  Gaps 0;

Qy        82 GGCGAGGGCCTCCGGCCACTACTATTCTTCTCAACATCCACCAGCATCAGCTCCTTCTCGC 141
             ||||||||||||||||||||| |||||| |||||  ||||| |||  ||| ||||||||||
Db       190 GGTGAAGGATTAAGACAATATTATTATTGCAACATATTCATGACGTTCAGCTTCAAGTCCGC 249

Qy       142 CAAAAGACTCATAACCTCAACCGTCCTCGAGGCTCAGAGAAACGAGACCTCAATTCTTCAAG 201
             |||||||||||| | |||||||||||||||||||||||||||||| |||||||||| ||||
Db       250 CAAAAGACTCATAATCTTAATCGTCTTGAGGCTCAACGTAACGAACTTAAGTTCTAGAGTA 309

Qy       202 AGGATGCTGGCGGAGAAAATTACAGCTTCTGCAGGAACCCGGCGTTTATGTCGGTGAAGTT 261
             | ||||||||||||||| || ||| |||||||||||||||||  |||| |||||||||||
Db       310 GGGATGCTTAAGGAAGAGTTGCGCCGCTGCTTCAAGAGCCCGGGTCTTATGTTGGTGAAGTT 369

Qy       262 GTCAAAGTAATGGGCAAGAACAAAGTCCTTGTCAAGGTCCACCAGCAAGGAAAATATGTT 321
             || |||||  |||||||| || |||||  ||| |||| ||||| || || ||||| ||||
Db       370 GTCAAAGTGATGGGCAAGTACTAGTCAAGGTGCACCCAGAGGGGTAAATATGTG 429

Qy       322 GTTGACATTGACAAAAATAATTGACATTACATTACTCCATCCACTAGAGAGTTGCACTC 381
             ||||||||||| |||||  ||||||| |  || |||||  || || ||| |||||||| |
Db       430 GTTGATATTGACAAAAGCATTGATATCACTAAGATAAACTCCATCAACCAGAGTAGCCCTC 489

Qy       382 CGCAACGACGAGTTATGTTCTTCACTAGTTCTGCCAAGTAAGTTGATCCTTGGTCAAT 441
             ||||||||||||||||||||||||  ||| |||  ||||| ||||  ||| || ||||
Db       490 CGCAATGACGACTACGTTCTTCATTGGTTTTGCCAAGAAGTTGATCCTTGGTCAAT 549

Qy       442 CTGATGAAAGTTGAGAAAGTTCCCGATTCTTACACATATGACCATGGTGCGTTTAGACCAG 501
             || ||||||||| || ||||| ||| |||| || ||||| || |||| ||| |||||||||
Db       550 CTCATGAAAGTCGAAAAAAGTTCCTGATTCTACTTATGGTGGGCCTTGACCAG 609

Qy       502 CAAATTAAAGAAAATAAAAGAGGGTCATTGAGCTACCACAAAACATCTGAGCTGTTTGAA 561
             ||||||||| ||| |  |||  ||||| ||| ||||||| ||| |||||||||||||||||
Db       610 CAAATCAGGAGATTAAGGAGGTTATTGAACTTCCAAATCAAAACATTCCAGAGTTATTCGAG 669

Qy       562 AGTCTTGGAATTGCACAACCACCAAAGGGTGTCCTTCTGCTCTATGGGCCACCTGGGTACA 621
             |||||||||| |||||||||||||||| || || ||| | ||||| ||||||| |||||||
Db       670 AGTCTTGGAATTGCTCAACCAAAAGGAGTTCTGTTATACGGACCTCCTGGTACCGGAAG 729

Qy       622 ACATTGTTGGCGTAGGGGCAGTGGCTCATCATCGACTGTACATGGGTGTCTCGGT 681
             ||||| ||||| ||||| | ||||||||||  ||||| || | ||||||||||||
Db       730 ACCTTGTTGGCGAGAGCAGTGGCTCATCATACTGATTGTGCCTTATCGGGTCTCGGT 789

Qy       682 TCTGAGTTAGTTCAGAAATACAATTGGAGAAGGTTCAGGAGACATTTTTGTT 741
             ||| || || |||||||| |  || || ||||||||||||||| |||||||||
Db       790 TCTGAGCTAGTGCAGAAATATATTGGAGAGGGCTCTCGGATGGTTCGGGAGCTTTTCGTC 849

Qy       742 ATGGCCAGGGAACATGCTCCATCAATTATCTTCATGGATGAAATTGACAGTATTGGATCT 801
             |||||||||||||||||| ||||||| | || ||||||||| |||||||| ||||| |||
Db       850 ATGGCCAGAGAACACATGCTCCATCATTTATTATGGATGAAATAGACAGTATCGGATCT 909

Qy       802 GCTCGGAATGGAAATCTGACAAGTGGCAAACCGGTAGTGAGTGGACGTACTACTGTCTGGAA 861
             ||||| |||||| | |||||| || || || ||||||| ||||| ||||| | ||||||
Db       910 GCTCGTATGGAGTCCGGAAGCGGAAATGGAGACAGTGAAGTGCAGAGAACTATGCTTGAG 969
```

```
Qy       862 CTTCTCAACCAGTTGGATGGATTCGAAGCTTCAAATAAGATCAAGGTTTTGATTGGCCACC 921
             ||||||| ||||||||||||||| |||||||||||||||| |||||||| |||| |||| |
Db       970 CTTCTTAACCAACTGGATGGATTTGAAAGCTCAACCAAAATCAAGGTTTTGATGGGCTACC 1029

Qy       922 AATCCGATTGGATATCCTGGATCACACCCCTCCTTAGACCAGCAGGACGATAGACCGGAAATT 981
             ||||||||||||||||||||||||||| |||  ||| || ||||| |||||| || ||||||
Db      1030 AACCGGATTGGATATCTTGGATCACACCCTTCCAGCCGATTGGACAGGAGAATTGATAGGGAGATT 1089

Qy       982 GAATTTCCAACCCTTAATGAAGAGTCTCGGCTGGATATTTTGAAAATCCATTCTAGAGA 1041
             || |||||||| || || |||||| || | ||||| || |||||| |||||| ||||||
Db      1090 GAATTCCAAACCCCAATGAAGAGATTCACCGTTGATATTCTGGATATTCCAAATTCAATTCAAGAAG 1149

Qy      1042 ATGAATTTAATGCGTGGCATTGAGTTGAGAAGATTGCCCGAGAAGAATGAATGCGGAGCATCT 1101
             |||| |||||||| ||| ||||| ||||||| ||| | | || |||| ||||| || ||| |
Db      1150 ATGAACTTAATGCGGGCGCATAGAGTTAAAGAAGAGTTGCTGAAGAGATGGAATGGCTCG 1209

Qy      1102 GGTGCCTGAACTTAAGGCGTGTTGTTGCACTGAAGCTGAAGTGTTGCTTTTGAGGCAGGAG 1161
             |||||||||||||||||||| ||||| ||||||||||||||||||||| |||||||||||
Db      1210 GGTGCTGAGCTTAAGGCTGTTTGTACTGAAGCTGGAATGTTTGCTTTGAGGGAAAGAAGAGA 1269

Qy      1162 GTACACGGTGACTCAGGAGGATATTTGAGATGGGCCGTTGGCGGAAGGTTGATGAAAAGGAGACT 1221
             ||| ||||| ||||||||||| ||||||| |||||||| | ||||| ||||| ||||||||||
Db      1270 GTCCATGTGACCCAAGAAGATTTTGAAATGGCAGTGGCGAGGTTATGAAGAAGGAAACT 1329

Qy      1222 GAAAAAAACATGTCATTGGGGAAGTTGTGGAAG 1254
             |||||| ||||| || ||||||| ||||| |||
Db      1330 GAGAAAAACATGTCTCTACGGCAAGCTATGGAAG 1362
```

```
BASE COUNT        436 a    309 c    407 g    364 t
ORIGIN

        ESGTGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIE
        FPNPNEDSRFDILKIHSRKMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRER
        RVHVTQEDFEMAVAKVMKDTEKNMSLRKLWK"

Qy  Query Match      64.7%;   Score 811.4;  DB 8;   Length 1516;
Db  Best Local Similarity   80.7%;   Pred. No. 2.6e-218;
    Matches  947;  Conservative   0;  Mismatches 226;  Indels   0;  Gaps   0;

Qy      82  GGCGAGGGCCTCCGCCACTACTATTCTTCTCAACATCCACGAGCTCCTTCTCTGGC  141
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138  GGGGAGGGGCCTCGCCGCAGTRACTACCTGCGAGCACACATCCACGAGATCCCG  197

Qy     142  CAAAAGACTCATAACCTCACCGTCTTCGCAGGCTCAGAGAAACAGACCTCAATTCTAGGGTG  201
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     198  CAGAAGAACGGCACACCTCAACCGCCTCGCAGGCCCAACGACCTCAACTCCAGGGTT  257

Qy     202  AGGGATGCTGCGGAAGAGAATTACACGCTTCTGCAGGAAACCCCGGCGTCTTATGTCGTGAAGTT  261
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     258  AGAATGCTCAGGGAGAGAGCTGGCAGTTGCTTCAAGGAACCTCGTATGTTGGTGAGGGTG  317

Qy     262  GTCAAAGTAATGGGCGAGAACAAAGTCCTTGTGTCAAGGTCCACCCAGAGGAAAATATGTT  321
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     318  GTGAAGGTCATGGGTAAAATCAAAGTTCTGAGTGGAAGGACGCAAAATATGTG  377

Qy     322  GTTGACATTGACAAAAATATTGACATTACAAGGAATTACTCCCACTAGAGTTGCACTC  381
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     378  GTGGATATTGATAAAAGCAATTGATTATTACAAAGGATCACACCTTCAACAAGAGTTGCTCTT  437

Qy     382  CGCAACGACAGTTATGTTCTTCACTAGTTCTGCCAAGTAAAGTTGAAGTTGATCCATTGGGTCAAT  441
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     438  CGAAAAGTAGCTATATGCTTCAATTGAATCCTGCCAAGCAAGAGTTGATCCGTTGGTCAAT  497

Qy     442  CTGATGAAAGTTGAGAAAGTTCCCGATTCCAGATTCTACATATGACATGATGTGSTGSTTTGACCAG  501
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     498  CTTATGAAAGTTGAGAAGGTTCCCGATTCTACTTATGATATGATTGGGAGGTCTTGAATCAG  557

Qy     502  CAAAATTAAAGACAAAATAAAAGAGGTCATTGAGCTACAAATAAAACATCCTGAGCGTGTTTGAA  561
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     558  CAAATCAAAGAGAGTCAAAGGAGTCATTGAACTTCCTATCAAACATCCAGAGCTATTTGAG  617

Qy     562  AGTTCTTGGGAATTGCACAACCAAAAAGGGTGTTCCTGCTCTATGGGGTACACGGGTAAA  621
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     618  AGTTCTTGGGAATTGCCCAACCAAAGGGTGTCCTTCTCTATGGGCCTTCACGGCCACAGGGAAAA  677

Qy     622  ACAATTGTTGGGCTAGGGCCAGTCCGAGTCTCATCATACTATACGGGTGTCTCGGT  681
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     678  ACATTCGTTGGCACGTGCAGTTGCTCATCACACTGACTGCCTTCATCAGGGTGTCTGGT  737

Qy     682  TCTGACTTAGTTCAGAAAATACATTGGAAGAAGGTTTCAGGGAACAGTTTTGTT  741
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     738  TCCGAGTTAGTTCAGAAGTACATTGGTAAGGTTCTCGGAATTGGTCCGATTGTCAAAACTT  797

Qy     742  ATGGCCAGGGAACATCTCCATCAATTATCTTAGACAGGATCAAAATTGCACAGTATTGGATCT  801
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     798  ATGGCCAGGGGGAACAATGCACCATCCATTATTTCATGGATGAAAATAGACTCCATTGGATCT  857

Qy     802  GCTCGGATGGAATCTGGAAGTCGGAGTGGGCAAACGTGACTTGATTGAGGTACAGGCGNACTATGCTGAA  861
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     858  GCTAGGATGGGAGTCTGGAACGGTGCAACGGTGACAGTGAACCATCGCTTGAG  917

Qy     862  CTTCTCAACCAGTTGCGGATGGATTTGAAGCTTCAAAATAAGATCCAAGGTGGTGGCCACC  921
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     918  CTTTCTGAACCAGTGATGTTTCGAGCATCAAACAAAAATAAGGTTCTGATGGCAACA  977

Qy     922  AATCGGATTCGATACTCCTGGATCAAGCCCTCTCCTTAGACCAGGACCGGATAGACCCGGAAAATT  981
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     978  AATAGGATAGAGCATCTTGGATCAGGCCCTTCGTCAGGCCTGGATAGACTGCAGGAAAAATT  1037

Qy     982  GAATTTCCAACCCTAATGAAGAGAGTCTCGGCGCTGGATATTTTTGAAAATCCATTCTAGAGA  1041
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1038  GAANTTTCCTAATCTAAATGAAGGAATTCACGCTTCAGGATCCACATCTTGAAGAGAA  1097
```

```
BASE COUNT        360 c    418 g    378 t                    469 a
ORIGIN

Qy  Query Match      64.7%;   Score 811.4;  DB 8;   Length 1625;
Db  Best Local Similarity   80.7%;   Pred. No. 2.7e-218;

Qy    1042  ATGAATTTAATCCGTGCATTGATTGATTTGCAAGAAGATTGCCGAGAAGATGAATGAGCATCT  1101
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1098  ATGAACTTGATGCGTGGTGGTATTGATCTGAAAAAGATCGCTGAAAAGATGAATGGGGCTTCA  1157

Qy    1102  GGTGCTGAACTTAAGGCTGTTTGCACTGAAGCTGGAATGTTTGCTTTGAGGGAGAGGCGGAGG  1161
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1158  GGAGCCGAGCTCAAGGCGGTGGTTATTGATCTGAACAGAACCTGGAATGTTTCGTCGTGGAAGG  1217

Qy    1162  GTACACGTGACTCAGGAGGAGGATTTGAGATGGCCGTGGCGAAGGTGATGAAAAAGGACACT  1221
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1218  GTGCCATGTAAACACAAGGAGGACTTTGAGATGGCAGTGGCGAAGGTGATGAAGAGAACACG  1277

Qy    1222  GAAAAAACATGTCATTGCGGAAGTTGTGAGAA  1254
            ||||||||||||||||||||||||||||||
Db    1278  GAAAAGACATGTCCCTACGGAAGGCTCTGGAAG  1310
```

RESULT 8
LOCUS   AX653352   1275 bp   DNA   linear   PAT 22-MAR-2003
DEFINITION Sequence 3222 from Patent WO03000898.
ACCESSION   AX653352
VERSION   AX653352.1 GI:29156166
KEYWORDS   .
SOURCE   Oryza sativa
  ORGANISM Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
            Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
  TITLE   Plant genes involved in defense against pathogens
  JOURNAL   Patent: WO 03000898-A 3222 03-JAN-2003;
            Syngenta Participations AG (CH)
FEATURES   Location/Qualifiers
     source   1..1275
                     /organism="Oryza sativa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4530"
BASE COUNT   358 a   267 c   350 g   298 t   2 others
ORIGIN

Query Match   64.6%; Score 810.6; DB 6; Length 1275;
Best Local Similarity 80.6%; Pred. No. 4.3e-218;
Matches 945; Conservative 2; Mismatches 226; Indels 0; Gaps 0;

Qy   82 GGCGAGGGCGCCTCCGCCACTACTATTCTCAACATCACCGAGCATCAGCTCCTTCTTCGC 141
            ||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   100 GGGGAGGGGCCTGCGGCAGTACTACCTGCACGCACATCCACGAGATCCGG 159

Qy   142 CAAAAGACTCATAACCTCAACCCGTCTCGGAGGCTCAGAGAAAACGACCTCAATTCTAGGGTG 201
            |||| |||||||||||||||||||||||||||||||| || |||||||||||||||||||||
Db   160 CAGAAGACGCACAACCTCAACCCGTCCCAGCGCAACGACCTCAACTCCCGAGTT 219

Qy   202 AGGGATGCTGCCGCGAAGAATTACAGCTTCTGCAGGAACCCGGCGCTCTTATGTCGGTGAAGTT 261
            ||||||||||||||||||||||||||||| || |||||||||||||||||||||||||||||||
Db   220 AGGATGCTGCCGCGAAGAGTCGCAGGAACCCGGCGTATGTTCGGTGAGGTG 279

Qy   262 GTCAAAGTAATGGGCAAGAACAAAAGTTCCTTGTCCAAGGTCACCCAGAAGGAAAATATATGTT 321
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   280 GTGAAGGTTCATGGGTAAATCAAAGGTTCCACCCAGAAGGCCAAAATATGTG 339

Qy   322 GTTGACATTGACAAAAATATTGACACATTACACAAAAGATTACTCCCATCCACTAGAGTTGCACTC 381
            ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   340 GTCGGATATTGATAAAAGCCATTGATATTACAAAGGTCCCTTGATATCCGGTGACTCAAT 399

Qy   382 CGCAACGGACAGTTATGTTCTTCACTTAGTTCTGTGCCAAGTAAAGTTGATCCATTGGTCAAT 441
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   400 CGCAACGACACAGTTATGTTCTTCACTTAGTTCTACTACAGTCTACAAGTAAGTTACTCAAT 459

Qy   442 CTGATGAAAGTTGGAAAGTTCCCGAATTCTACATATGACATGATTGGTGGGTTTAGACCAG 501
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   460 CTTATGAAAGTTGAGAAGGTTCCTGATTCTACTTATGATCTCGAGAGGTCTTGATCAG 519

Qy   502 CAAATTAAAGAAAATAAAAGAGGTCATTGAGCTACCAATCAAACATCCTGAGCTGTTTGAA 561
            |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   520 CAAATCAAAGAGAATAAAAGAGGTCATTGAACTTCCTATCAAACATCCAGAGCTATTTGAG 579

Qy   562 AGTCTTGGAATTGCACAACCAAAGGGTGTTCCTGCTCTATGGCCACCTGGTACTAGCTAA 621
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   562 AGTCTTGGAATTGCACAACCAAAGGGTGTTCCTGCTCTATGGGCCACCTGGTACAGGTAA 621

Qy   1102 GGTGCTTGAACTTAAGGGCTGTTTGCACTGAAGCTGGAAATGTTTGCTTTGAGGAGCGGGAGG 1161
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   520 CAAATCAAAGAGAATAAAAGAGGTCATTGAACTTCCTATCAAACATCCAGAGCTATTTGAG 579

Qy   1269 GGAGCCGAGCTCAAGGCGGTCTGCACAGAGCTCGGAAATGTTTTGCTCTTCGGTGAGAGAAGG 1328
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   562 AGTCTTGGAATTGCACAACCAAAGGGTGTTCCTGCTCTATGGGCCACCTGGTACAGGTAA 621

RESULT 9

LOCUS   AY079403   1291 bp   mRNA   linear   PLN 18-SEP-2002
DEFINITION Arabidopsis thaliana putative 26S proteasome AAA-ATPase subunit
            RPT6a (At5g20000) mRNA, complete cds.
ACCESSION   AY079403
VERSION   AY079403.1 GI:19310806
KEYWORDS   FLI_cDNA.
SOURCE   Arabidopsis thaliana (thale cress)
  ORGANISM Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 1291)
  AUTHORS   Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Etgu,P.,
            Lee,J.M., Toriumi,M., Yu,G., Brooks,S., Chao,Q., Chen,H.,
            Karlin-Neumann,G., Kim,C., Lam,B., Miranda,M., Nguyen,M.,
            Palm,C.J., Shinn,P., Southwick,A., Davis,R.W., Ecker,J.R. and
            Theologis,A.
  TITLE   Arabidopsis Open Reading Frame (ORF) Clones
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 1291)
  AUTHORS   Yamada,K., Banh,J., Chan,M.M., Chang,E., Dale,J.M.,
            Deng,J.M., Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L.,
            Tang,C.C., Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L.,

COMMENT   Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
            Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
            Lin,J., Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J.,
            Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A.,
            Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
            Direct Submission
            Submitted (19-FEB-2002) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
            The RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA ; 'RIKEN
            Arabidopsis Full-length (cDNA)'; Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members constructed and
            sequenced the pUNI (ORF) clones using the RAFL cDNAs: Yamada,K.,
            Banh,J., Chan,M.M., Chang,C.H., Chang,E., Dale,J.M., Deng,J.M.,
            Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C.,
            Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L., Chen,H.,
            Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Lam,B., Lin,J.,
            Meyers,M.C., Miranda,M., Nguyen,M., Palm,C.J., Shinn,P.,
            Southwick,A., Davis,R.W., Ecker,J.R. and Theologis,A.

            Yamada,K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally
            to this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP
            /PGEC) contributed equally to this work as PIs.

            Annotation is based on the January 2002 version of the Arabidopsis
            genome submitted to GenBank.
FEATURES   Location/Qualifiers
     source   1..1291
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="U10022"
                     /note="This clone is in pUNI 51.
                     ecotype: Columbia"
     gene   1..1291
                     /gene="At5g20000"
     gene   1..1291
                     /gene="At5g20000"
     CDS   1..1260
                     /gene="At5g20000"
                     /codon_start=1
                     /evidence=experimental
                     /product="putative 26S proteasome AAA-ATPase subunit
                     RPT6a"
                     /protein_id="AAL85134.1"
                     /db_xref="GI:19310807"
                     /translation="MAAVGVEARPPVTAMEETCNVKGAAAKQGEGLNKYYLQHLDELQ
                     RLQREKSYNLRLEAQRNELNSRVRMLREELQLLQEPGSVGEVKVMGKNKVLVKVH
                     PEGKYVVDIDKSIDITKLTPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVDSTY
                     DMIGGLDQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTILARAVAHH
                     TDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGSG
                     NGDSEVQRTMLELNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPNPN
                     EESRFDILKIHSRKMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHVT
                     QEDFEMAVAKVMKKDTEKNMSLRKLWK"
     3'UTR   1261..1291
                     /gene="At5g20000"
BASE COUNT   383 a   235 c   348 g   325 t
ORIGIN

Query Match   64.6%; Score 810; DB 8; Length 1291;
Best Local Similarity   80.2%;
Matches 951; Conservative   0; Mismatches 235; Indels   0; Gaps   0;

Qy   69 GCCCACCAAGCAGGGGCGGAGGGGCCTCCGCCACTACTATTCTTCAACATCCACGACATCA 128
               ||| || |||||| |||| |||||||||||||||||||||||||||||||||||||||||||
Db   72 GGCGGCGAAACAAGGAGGAGAGGGGCTTAACAAGTAACTACCTCCAGCACTCGATGAGCTCCA 131

Qy   129 GCTCCTTCTTCGGCCAAAAGACTTCATAACCTCAACCGTCTCGAGGCTCAGAGAAACGACCT 188
               || ||||||| || |||| |||||||| | ||||||| |||||| |||||||||| ||
Db   132 GCGCCTGCAACGGCGAAAAGTCGTATAACCTTAATCGTCTTGAAGCTCAGAGGAATGAACT 191

```
RESULT 10
LOCUS       AY087503                1475 bp    mRNA    linear   PLN 14-APR-2003
DEFINITION  Arabidopsis thaliana clone 36085 mRNA, complete sequence.
ACCESSION   AY087503
VERSION     AY087503.1  GI:21406240
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1475)
  AUTHORS   Haas,B.J., Volfovsky,N., Town,C.D., Troukhan,M., Alexandrov,N.,
            Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.
  TITLE     Full-length messenger RNA sequences greatly improve genome
            annotation
  JOURNAL   Genome Biol. 3 (6), RESEARCH0029 (2002)
  MEDLINE   22088475
   PUBMED   12093376
REFERENCE   2  (bases 1 to 1475)
  AUTHORS   Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
            Feldmann,K.
  TITLE     Full-length cDNA from Arabidopsis thaliana
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1475)
  AUTHORS   Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
            Feldmann,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,
            Malibu, CA 90265, USA
COMMENT     This clone sequence is one of 5,000 Ceres full-length cDNAs made
            available to TIGR and Genbank. The following quality assessment of
            this set was done by comparison with known proteins: two percent of
            the clones are estimated to be 5'-truncated; less than one percent
            are 3'-truncated; approximately two percent represent alternative
            splice variants, including unspliced introns and spliced exons; one
            percent may contain premature stop codons; five percent may have
            frame shifts in a coding region. A sequence is considered to be
            5'-truncated if it lacks the translation initiation start (ATG). A
            sequence is considered to be 3'-truncated if it lacks the
            C-terminal end of the encoded protein. Please note that these cDNA
            sequences are derived from the Ws or Ler ecotypes and therefore
            may contain polymorphisms when compared to sequences from Col-0.
            Genset carried out the library production and sequencing of the
            full-length clones. Ceres, Inc. carried out the clustering of the
            5' sequences, selection of clones, and sequence assembly.
FEATURES             Location/Qualifiers
     source          1..1475
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
     CDS             75..1334
                     /codon_start=1
                     /product="26S proteasome AAA-ATPase subunit RPT6a-like
                     protein"
                     /protein_id="AAM65046.1"
                     /db_xref="GI:21593097"
                     /translation="MAAVGVEARPPVTAMESTCNVKGAAAKGGEGGLNKYVLQHLDELQ
                     RLQREKSYNLNRLEAQRNELNSRVRMLREELQLLQEPGSYGVKVMGKNKVLVKVH
                     PEGKYVVDIDKSIDITKLIPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTY
                     DMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHH
                     TDCTPIRVSGSSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGSG
                     NGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPNPN
                     ESRFDLIKIHSRKMNIMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHVT
                     QBDFEMAVAKVMKKDTEKNMSLRKLMK"
BASE COUNT      424 a    271 c    378 g    402 t
ORIGIN

Query Match     64.6%;  Score 810;  DB 8;  Length 1475;
Best Local Similarity   80.2%;  Pred. No. 6.5e-218;
Matches  951;  Conservative  0;  Mismatches  235;  Indels  0;  Gaps  0;
```

```
Qy    189 CAATTCTAGGTGAGGATGCTGCGGAAGAATTACAGCTTCGCAGGAACCCGGCGTCCTTA 248
          |||||| || | | |||||| | | ||||| |||||||||    |  ||||||| | |||||
Db    192 CAATTCTGAGTACGAATGCTCAGAGAAGAGTTACAGCTTCCTTCCTTCAAGAACCTGGGGTCCTTA 251

Qy    249 TGTCGGGTGAAGTGTTCAAGTAATGGCGTCGGAGAAGAGTTACAGCTTCCTTCCTCCAGA 308
          |  | ||||||||| |  | ||| |||||||| |||||| ||||| || || |||||||
Db    252 TGTGGGGTGAAGTGGTAAAAGTGATGGGGAAGAACAAGGTTCTTGTCAAGGTTCATCCGAG 311

Qy    309 AGGGAAAATATTGTTGGTTGACATTGACAAAAATAATTACAAAGAATTACTCCATCCAC 368
          || ||||   |||||| | ||||| | |||| ||| |||||  |||| ||||||||||
Db    312 GGGGAAGTATGTTGTTGTCGATATTGACAAAAGTATAGACATAAAACTCACTCCATCAAC 371

Qy    369 TAGAGTTGCACTCCGCAACGACGAGTTATGTTCTTCACTTAGTTCTGCCAAGTAAAGTTGA 428
          ||| ||| ||| |||| |  ||  || | ||| || |||| ||| || ||  |||| ||
Db    372 GAGAGTTGCTCTCCGTAATGATAGTTATGTTCTTCACCTGGTTCTGCCAAGTAAAGTAGA 431

Qy    429 TCCATTGGTCAATCTGATGAAAGTTGAGAAAGTTCCGGATTCTTACATATGACATGATTGG 488
          || ||| |||| || |||  ||| || |||| |||| ||||| || |  || ||||||||||
Db    432 TCCCTTGGTTAACCTGATGAAAGTTGAGAAAGTTCCGGATTCTACCTATGATGATGATTGG 491

Qy    489 TGGTTTAGACCAGCAAATTAAAGAAATAAAGAGGTCATTGAGCTACCAATCAAACATCC 548
          ||||| || || | |||| |||| || || || || ||| ||||||| | || ||||
Db    492 TGGTCTTGACCAGCAGATTAAAGAAATAAAGGAGGTCATTGAACTGCCAATCAAGCATCC 551

Qy    549 TGAGCTGTTTGAAAGTTCTTGGAATTGACACACCAAAGGGGTTCTCCTGCTCTATGGGCCACC 608
          |||| || |||| |||   |||  || | | |||||| ||||| || ||| || || ||||||
Db    552 TGAGCTGTTTGAAAGTCTTGGAATTGCACAACCAAAGGGTGTTCTTCTCTATGGTCCACC 611

Qy    609 TGGTTACAGGTGAAAACATTGTTGGCTAGGCAGTGGCTCATCATACTGACTCTACAATTCAT 668
          |||| |  | ||||| |||| || ||| || ||| ||| ||||||||| ||| ||  ||||||
Db    612 TGGAACTGGGAAAACACTATTGGCTAGGGCAGTGGCTCATCACACTGACTGTACTCCTATCAT 671

Qy    669 CAGGGTGTTCGGTTCGTTCGAGTTAGTTCAGAAATCATATTGGAGAAGAGTTCTAGAATGGTCAG 728
          ||||| ||| ||| | ||| |||| |||||||  || ||| |||||||||||| ||||||| |||
Db    672 CAGGGTTTCTGGTTCCAGAAATACATTGGAGGAAGTTCTAGAAGGTTCTAGAAATGGTCAG 731

Qy    729 GGAACTTTTTGTTATGGGCCAGGGGAAACAGCTCCATCAATAATTACTTCATGGATGAAAATTGA 788
          || | ||||| || || ||||| |||||| ||||||||| |||| || |||| ||||| | |||
Db    732 AGAACTTTTTGTGATGGCGAGGGAACACGCTCCATCAATCATCTTCATGGATGAGATTGA 791

Qy    789 CAGTATTGGATCTCGCTCCAACCAGTGATGATTGGAATCTGGTAGTAGTTGGAGGTAGCGAGAATCGA 848
          |||  ||||| |  |||| ||| ||| | || ||| |  ||| ||| | ||| || |   ||||
Db    792 TAGTATCGGGTCTGCTCGTATGGAAAGTGGGTCAGGGAATGGTGACAGTGAGGTTCAGCGAACTATG 851

Qy    849 GACTAATGCTGAGCTTCTCAATCAACTTGACGGAGGTCAAACAAAATCAAGGT 908
          ||| | | || |||| ||||| ||||| ||||| || |  |  |||||||||
Db    852 GAGCTTCTCAACCAGCTTGACGGGTTTGAAGCTTCTAATAAGATCAAGGT 911

Qy    909 TTTGATGGCCACCAATCGGATCGATATCCTTGGATCAAGACCCAGGAGGGAT 968
          |||| ||| | |||| || ||| | | ||| || || || | |||| || |||
Db    912 CCTTATGGCCACGAATCGTATTGATATCCTTGACCAAGCTCTCCTGCGACCAGGAAGGAT 971

Qy    969 AGACCCGGAGAAATTGAAATTTCCAACACCCTAAGTGAAGGAGTCTCGGCGGCTGGGAATATTTTGAAAAT 1028
          ||| | |||||| || || |||| |  |||  || || | || ||| || ||  || | ||||| ||||
Db    972 TGATAGGAAAATTGAATTTCCTAATCCTAATGAGTCACGTTTTGATCTTATCAAAATTCATTCAAGAAAAT 1031

Qy   1029 CCATTCTAGAAGAATGAATTAATCGCTGGCATTGATTTGAAGAAGATTGCCGAGAAGAT 1088
          || | | |||| || || || ||  | ||||| || |||||||| ||| |||||||| |
Db   1032 ACACTCAAGGAAATGAATATTATGCGTGGCATTGATTTGAAGAAGATTGCTGAGAAGAT 1091

Qy   1089 GAATGGAGCATCTGGTGCTGAACTTAAGGCTGTTTGCACTGAGGCTGGAATGTTTGCTTT 1148
          ||| || || ||||||| ||| ||||| ||||| | || ||| || |||||||| ||||
Db   1092 GAATGGTGCTTCAGGTGCTGAGCTGAAGGCAGTTTGCACTGAAGCTGGCATGTTTGCTCT 1151

Qy   1149 GAGGGAGCGGCGGGTTCACGTGACTCAGGACGATTTTGAGATGGCCGTGGCCAAGGTGAT 1208
          |||| || ||||| || |||||| |||| || |||||| ||||| ||| | ||||| ||
Db   1152 TCGTGAGAGAAGAGTACATGTGACTCAGGAAGATTTTGAGATGGCTGTTGCCAAGGTAAT 1211

Qy   1209 GAAAAGAGAGACTGAAAAAAACATGTCATTCGGGAGTAAGCTCTGGAAG 1254
          ||||| |  || | |||| |||| ||| | | || | || || || |||
Db   1212 GAAGAAAGACACAGAGAAGAACATGTCTCTGCGGAAGCTCATGAAGTAAT 1257
```

```
Qy     69 GCCCACACAAGCAGGGGCGAGGGCCTCGCGACTATTCTCTGAACATCCACGAGCATCA  128
Db    266 CAATTCTCGAGTACGAAGATGCTCAGAGGAGGAGTTACGACCTTCAAGAACCTGGGTCCTA 325
Qy    146 GGCGGGCCGGAAACAAAGGAGGAGGGCTTAACAAGTACTCATCCTCCCAGCAATCTCGATC 205
Db    206 GGCCCTGCAACCGCAAAAGTGGTATAAACCTTAATCGTTCTTGAAGTCTCAGAGGAATGAACT 265
Qy    129 GCTCCCTCCTTCGCGAAAAAGGACTCATAACCGTCCAGGCTCGAGGCTCAGAGAACGACCT 188
Db    249 TGTCGGCTGAAGTTGTCAAAGTAATGGGCCAAGAGTAATCTCCACCAGA 308
Qy    206 GGCCCTGCAACCGCAAAACGGTATAAACCTTAATCGTTCTTGAAGTCTCAGAGGAATGAACT 265
Qy    189 AGGAAAATATGTTGTTGACATTGACAAAATATTGACATTACAAAAGATTACTCCATCCAC 248
Db    266 CAATTCTCGAGTACGAAGATGCTCAGAGGAGGAGTTACGACCTTCAAGAACCTGGGTCCTA 325
Qy    309 AGGAAAATATGTTGTTGACATTGACAAAATATTGACATTACAAAAGATTACTCCATCCAC 368
Db    386 GGGGGAGTATGTTGTCGATAATTGACAGAATGATTGAACATAACCAGTAAGGTCCATCAAC 445
Qy    326 TGTGGGTGAAGTTGGTAAAAGTGATGGGAAAAACAAGGGTCTTGGTTAAGGTTCATCCAGA 385
Db    369 TAGAGTTGCACTCCGGAACGACAGTTATGTTCTTCACTTAGTTCTGCCAAGTAAAGTTGA 428
Qy    386 GGGGGAGTATGTTGTCGATAATTGACAGAATGATTGAACATAACCAGTAAGGTCCATCAAC 445
Db    429 TCCATTGGTCAATCTGATGAAAGTTGAGAAAGTTCCGATTCTACATATATGACATGATTGG 488
Qy    429 TCCATTGGTCAATCTGATGAAAGTTGAGAAAGTTCCGATTCTACATATATGACATGATTGG 488
Db    506 TCCCTTGGTTAACCTTATGAAAGTTGAGAAAGTTCCAGACTCCACATATGACATGATTGG 565
Qy    446 GAGAGTTGCCTCTCCGGTAATGATAGCTAATGTTCTCCACCTGGTTCTGCCAAGTAAAGTAGA 505
Db    506 TCCCTTGGTTAACCTTATGAAAGTTGAGAAAGTTCCAGACTCCACATATGACATGATTGG 565
Qy    489 TGGGTTTAGACCAGCAGCAAATTAAAGAAAATTAAAGAGAGTCAATGAGTCATTGAGAAATC 548
Db    566 TGGTCTTGACCAGCAAATCAACAGAAATAAAGAGAGTCCATTGAACATCTCCATCCC 625
Qy    549 TGAGCTGTTTGAAAGTCTTGGAAATTCCAACCAAAAGGGTGTCCTTATTGGGCCACC 608
Db    626 TGAATTGTTTGAGTCTCTTGGAAATTCCGCGCAAAAGGGTGTCGTTGTTGTACGGTCCACC 685
Qy    609 TGGTACAGGTAAAACATTGTTCGGCTAGGGACAGTGGCTCATCATACTGACTACATTCAT 668
Db    686 TGGAACTGGGGAAGACAACTATTGGCTCGGGTCGGTCGGGCTCATCATACGTACTTTCAT 745
Qy    669 CAGGGTGTCTGGTTCTGAGTTGTAGTTCAGAAGAAGGGTTCTTAGAATGGTCAG 728
Db    746 CAGAGTTTCTGGTTCTCGAGCTCCAGAAAATACATTGGAGGCAGGTTCTAGGAATGGTCAG 805
Qy    729 GGAACTTTTTGTTATGGCCAGGGAACATGCTCCATCCATTATATCTTCCATGGATGAAATTGA 788
Db    806 AGAACTTTTTGTTGATGGCGACCAACATCATCTTCCATGGATGAAAATCGA 865
Qy    789 CAGTATTGGGAATGCTCGTCGGAAATCTGGAAGTGGCGAACGGTGATAGTGAGGTACAGCG 848
Db    866 TAGTATGTCGGGTGTCGCTCGGTATGGGAATCTGGTTAGTGGAAATGGTGGGAAATGGTGGCGAAG 925
Qy    849 TACTATGGCTGGGAACTTCTCAACCAGTTGGATGGGATTTGAAGCTTCAAATAAGATCAAGGT 908
Db    926 GACTACTCCTTGAGCTCTCACACTCAACTTACGGGATTCGAGGCGTGCAAACAAAATCAAGGT 985
Qy    909 TTTGATAGGCCACCAATGGATTGGATGATATCGGAATGGATAAAGAGTTCTGGAATTCCTGGAAGGAT 968
Db    986 TTTGATGGCTACAAAATGTATTGATAATTCGGATCAAGCTCGATTCTGGATCAGGAATCCTGGAAGGAT 1045
Qy    969 AGACCGGAAAATGAATTCAAACCCCTAATGAAGAAGAGTCTCGGCTGCGATATTTTGAAAAT 1028
Db   1046 TGATAGGAAAATGAATTGAATTTCCTAATCTGAAAGAGTCACGGTTTAATCTTGAGATCTTGAGAT 1105
Qy   1046 CCATTCTGAGAAGAAGATTAATTCCTAATCTGAAAGAGTCACGGTTTAATCTTGAGAT 1105
Db   1106 ACACTCAAGGAAAATGAATTGATGGCTGAATCGATCTGAAAAAAGAGAATTGCAGAGAAGAT 1165
Qy   1089 GAATGAGAGCATCTGGTGCTGAACTTAAGGCTGTTTGCACTGAAAGCTGGAATGTGTTTTGCTTT 1148
Db   1106 ACACTCAAGGAAAATGAATTGATGGCTGAATCGATCTGAAAAAAGAGAATTGCAGAGAAGAT 1165
Qy   1149 GAGGGGACGGGAGGGGTACACGTGACTGACTCAGGAGGAGATTTTGAGATTGGGGCCCGTCGTGGCGGAAGGTGAT 1208
Db   1226 TCGTGAGAGGAGTACATGTGACTCAGGAGAACTTGGAGATTGGGCCGTGCTGTGGCGGAAGGTAAT 1285
Qy   1209 GAAAAAGGAGACTGAAAAAACACATGTCATTGCGGAAGTTGCGGAAG 1254
Db   1286 GAAGAAAGACACAGAGAAGAAACATGTCTCTGCGGTAAGCTCTGGAAG 1331
Qy   1166 GAATAGGTGCTTCAGGTGCTGACCTGAAGGCTGTATGCACGGAGGGGGCCAGGTTTTGCACT 1225
```

RESULT 11

AY039965

LOCUS   AY039965   1512 bp   mRNA   linear   PLN 18-SEP-2002
DEFINITION Arabidopsis thaliana putative 26S proteasome AAA-ATPase subunit
            RPT6a (At5g20000) mRNA, complete cds.
ACCESSION   AY039965
VERSION   AY039965.1 GI:14532681
KEYWORDS   FLI_cDNA.
SOURCE   Arabidopsis thaliana (thale cress)
  ORGANISM Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 1512)
  AUTHORS   Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
            Dale,J.M., Gibson,H.A., Goldsmith,A.D., Jiang,P.X., Lee,J.M.,
            Quach,H.L., Tang,C.C., Toriumi,M., Yu,G., Bowser,L., Carninci,P.,
            Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
            Karlin-Neumann,G., Kawai,J., Kim,C., Koesema,E., Lam,B., Lin,J.,
            Lin,J., Miranda,A., Narusaka,M., Nguyen,M., Palm,C.J.,
            Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J.,
            Satou,M., Seki,M., Shinn,P., Southwick,A., Sakurai,T.,
            Shinozaki,K., Davis,R.W. and Theologis,A.
            Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE   Arabidopsis Full Length cDNA Clones
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 1512)
  AUTHORS   Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
            Dale,J.M., Gibson,H.A., Goldsmith,A.D., Jiang,P.X., Lee,J.M.,
            Quach,H.L., Tang,C.C., Toriumi,M., Yu,G., Bowser,L., Carninci,P.,
            Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
            Karlin-Neumann,G., Kawai,J., Koesema,E., Lam,B., Lin,J.,
            Meyers,M.C., Narusaka,M., Nguyen,M., Palm,C.J.,
            Satou,M., Seki,M., Shinn,P., Southwick,A., Tracy,S.E.,
            Miranda,M., Nguyen,M., Palm,C.J., Shinn,P., Southwick,A.,
            Tracy,S.E., Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE   Direct Submission
  JOURNAL   Submitted (11-JUN-2001) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
COMMENT   RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA'; Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Yamada,K., Liu,S.X.,
            Sakano,H., Pham,P.K., Banh,J., Chung,M.K., Dale,J.M., Gibson,H.A.,
            Goldsmith,A.D., Jiang,P.X., Lee,J.M., Quach,H.L., Tang,C.C.,
            Toriumi,M., Yu,G., Bowser,L., Chen,H., Cheuk,R., Jones,T.,
            Karlin-Neumann,G., Kim,C., Koesema,E., Lam,B., Lin,J., Meyers,M.C.,
            Miranda,M., Nguyen,M., Palm,C.J., Shinn,P., Southwick,A.,
            Tracy,S.E., Davis,R.W., Ecker,J.R. and Theologis,A.

            Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed equally to
            this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGEC)
            contributed equally to this work as PIs.

            Annotation is based on the January 2002 version of the Arabidopsis
            genome submitted to GenBank.
FEATURES   Location/Qualifiers
     source   1..1512
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="RAFL05-07-D23 (R10022)"
                     /note="This clone is in a modified pBluescript vector

```
          (Lambda ZAP) as a XhoI/SstI insert.
          ecotype: Columbia"

gene      1..1512
          /gene="At5g20000"

5'UTR     1..78
          /gene="At5g20000"

CDS       79..1338
          /gene="At5g20000"
          /codon_start=1
          /product="putative 26S proteasome AAA-ATPase subunit
          RPT6a"
          /evidence=experimental
          /protein_id="AAK64142.1"
          /db_xref="GI:14532682"
          /translation="MAAVGVEARPVTAMEETCNVKGAAAKQGEGLNKYYLQHLDELQ
          RLQREKSYNLNRLEAQRNELNSRVRMLREELQLLQGEPGSYVGEVKVMGKNKVLVKVH
          PEGKYVVDIDKSIDITKLTPSTRVALRNDSYVLHIVLPSKVDPLVNIMKVEKVPDSTY
          DMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHH
          TDCTFIRVSSSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGSG
          NGDSEVQRTMLELMNQLDGFEASNIKVLMATNRIDILDQALLRPGRIDRKIEFPNPN
          EESRFDILKIHSRKMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHVT
          QEDFEMAVAXVMKXDTEKNMSLRKLMK"

3'UTR     1339..1512
          /gene="At5g20000"

misc_difference  1496
          /gene="At5g20000"
          /note="not present in genomic sequence"

BASE COUNT    446 a    278 c    379 g    409 t
ORIGIN
```

```
Qy     69 GCCACCAAGCACAGGGGCCGAGGGGCCTCCGCCACTACATTCTCTCAACATCCACGAGCATCA 128
          ||||||| |||||||||||||||||||||||| |||||||||  |||||||||||||||||
Db    150 GGCGGGCGGAAACAAGGAGAGAAGGGCTTAACAAGTACTACCTCCAAGCATCTCGATGAGCTCCA 209

Qy    129 GCTCTTCCTTCTCCGCCAAAAGACTCAATAACCCTCCTCTCGAGGAAAACGACCT 188
          |||||   |||||||   ||||||||||||||||| |||| |||||||||
Db    210 GCGGCCTTGCAACCGGAAAGTCGTATAACCTTAATCGTCTTGAAGCTTCAGAGGAATGAACT 269

Qy    189 CAATTCTTAGGGTGAGGAATGCTGCGCGAAGAATTACGCTTCTGCGCAGGAACCGGCCTCTTA 248
          |||||||  ||||||||||||||||||||||||||||||||||||| |||||
Db    270 CAATTCTTCGAGTACGAATTGCTTCAGAGGAGTTACAGCCTCCTCCTCAAGAACCTGGGTCCTA 329

Qy    249 TGTCCGGTGAAGTTGTCAAAGTAAATGGGCCAAGAACAAAAGTCCTTGTCAAGGTCCACCCAGA 308
          |||||||||||||||||||||||  |||||||||||| ||||||| |||||||||||||||
Db    330 TGTCGGGTGAAGTGGTAAAAGTGATGGGCCAAAAACAAGGTCTGGGTTAAGGTTCATCCAGA 389

Qy    309 AGGAAAAATATGTTGTTGACATTGACAAAAATATTGACATTACAAAGATTACTCCATCCAC 368
          |||||||||||||||||||||| ||||||||||||||||| |||||||||||||
Db    390 GGGGAAGTATGTGTTGCATATTGACAAAAGTATAGACATTACGAAACTCACTCCATCAAC 449

Qy    369 TAGAGTTGCACTCCGCAACGACACAGTTATGTTCTTCACTAGTTCTGCCAAGTAAAGTTGA 428
          ||||||||| ||||| |||||||||| ||| ||||||||| |||||||||||||||||
Db    450 GAGAGTTGCTCTCCGTAATGATAGCTATGTTCCCAAGTCTGGTTCGCCAAGTAAAGTAGA 509

Qy    429 TCCATTGGTCAATCTCGATGAAGATGTGAGAAAGTTCCCGGATTCTACATATGACATGATTCGG 488
          ||||||||||  ||||||||||| ||||||||| ||||||||||||||||||||| |||||
Db    510 TCCCTTGGTTAACCTTATGAAAGTTGAAAAGGTTCCAGACTCACCATATGACATGAITTGG 569

Qy    489 TGGTTAGACCAGCAAATTAAAGAAATAAAGAGGTCATCAITGAGCTACCAAITCAAACATCC 548
          |||||||||||| ||||||||||  |||||||||||||| |||||||||||| |||||||
Db    570 TGGGTCTTGACCAGCAAATCAAGGAAATAAAGGAGTCATTGAGTAGCTACCAATCAAAGCAICC 629

Qy    549 TGAGGCTGTTTGAAAGTCTTGGAATTGCACAACCAAAGGGTGTTCTGCTCTAGGCCACC 608
          |||||||||||||||||||  ||||||||||| |||||||| ||||||||| ||||||
Db    630 TGAATTGTTTGAGTGTCTTCGGAATTGCGCAGCAAAGGGTTGTTTGTTGTTGGTACGGTACC 689

Qy    609 TGGTACAGTAAAACATTGTTGGGTPAGGGCAGCAGCAGGTGGCTCATCATACTTGACTTTCAT 668
          || |||||||||  ||||||||||||||||||| |||| |||||||||||||||||||||
Db    690 TGGAACTCGGAAGAGACACTATTGGCTCGGCGCACATCACTGACTGACTTTCAT 749
```

Query Match   64.6%;   Score 810;   DB 8;   Length 1512;
Best Local Similarity 80.2%;   Pred. No. 6.6e-218;
Matches 951;   Conservative 0;   Mismatches 235;   Indels 0;   Gaps 0;

```
RESULT 12
AF220199
LOCUS       AF220199                  1747 bp    mRNA    linear    PLN 26-JAN-2000
DEFINITION  Pinus taeda 26S proteasome regulatory subunit 8 mRNA, complete cds.
ACCESSION   AF220199
VERSION     AF220199.1  GI:6752879
KEYWORDS    .
SOURCE      Pinus taeda (loblolly pine)
  ORGANISM  Pinus taeda
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE   1  (bases 1 to 1747)
  AUTHORS   Destefano-Beltran,L.J.C., Casas-Mollano,A. and Cairney,J.
  TITLE     Isolation of a full-length cDNA encoding a putative 26S proteasome
            subunit 8 from Pinus taeda somatic embryos
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1747)
  AUTHORS   Destefano-Beltran,L.J.C., Casas-Mollano,A. and Cairney,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Forest Biology, Institute of Paper Science
            and Tecnology, 500 10th St. NW, Atlanta, GA 30318, USA
FEATURES             Location/Qualifiers
     source          1..1747
                     /organism="Pinus taeda"
                     /mol_type="mRNA"
                     /isolate="260"
                     /db_xref="taxon:3352"
                     /clone="pLPSE7.1"
                     /tissue_type="somatic embryo"
     CDS             38..1339
                     /note="similar to Tat-binding protein; AAA-ATPase family
```

```
Qy    669 CAGGGGTGTCTGGGTTCTCGAGTTGAGTTCAGAAATACAITAGGAGGAGGGTTCTTAGAATGGTCAG 728
          |||||||||||||||||||||||||||||||||||| |||||||| |||||||| ||||||||||
Db   1088 CAGAGTTTCTCGGTTCTGAGCGTGGTCCAGAAATACAITAGGAGAAGGGTTCTAGAATGGTTCAG 728

Qy    729 GGAACTTTTTGTTTGTATGCCAGGGAACATGCTTCCATCAAITTATCTTCATGGATGAAAATTGA 788
          ||||||||||||||| |||||||||||||||||||| |||||||||| ||||||||| |||||
Db   1109 CAGAGTTTCTCGGTTCTGAGCGTGGTCCAGAAATACAITAGGAGAAGGGTTCTAGAATGGTTCAG 809

Qy    789 CAGTATTGGATCTGCTCGGAGTGGAAATCTGGAAGTGGCAACGGTGATAGTGAGGTACACG 848
          ||||||||||||| |||||||||||||||||||| ||||||||||| |||| |||||||
Db   1029 CCAITCTTAGGAAGAATGAAATTTAAATGGGCTGGGGGCCATTGAAGACTGAAGATGGTACACG 1088

Qy    849 TACTATGCTGGAACTTCTCAACCAGTGGGATGGATTGAAGCTTCAAATAGATCAAGGT 908
          ||||||||||||||||||||||| |||||| ||||||||||| |||||||| |||||||
Db   1050 TGATAGGGAAAATTGAAITTCCTAATCCTAATGCAGAAGACTGAAAGACGTTTGATATCTTGAAGAT 1109

Qy    909 TTTGATGGCCACCAATCGGATTGATATCCTGCATCAAGCCCTCCTTAGACCAGGACGGAT 968
          |||||| ||||||||| ||| |||||||||||||||||||||||| |||||||||||||
Db    969 AGACCGAAAAATTGAATTTCCAACCCCTAATGAAGACGTCTCGGCTCGGATTATTTGAAAAT 1028
```

```
          member"
          /codon_start=1
          /product="26S proteasome regulatory subunit 8"
          /protein_id="AAF27916.1"
          /db_xref="GI:6752880"
          /translation="MRVSELGGILKPASDLEAANAKEKLKEKHQLRESNRMVKHGE
          GLRQYYLHIHDLQLQVRQKTHNILNRLEAQRNDLNSKVRALREEIQLLQEGSYVGEV
          VKVMGKSKVLVKHPEGKYIVVDVDXNIDITKITPSTRVALRNDSYLHLLLPSKVDPL
          VNLMVEKVPDSTYDMIGGLDDQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPP
          GTGKTLLARAVAHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDE
          IDSIGSARMESGTGNGDSEVORTMLELLNQLDGFEASNIKIVLMATNRIDILDQALLR
          PGRIDRKIEFPNPNEESRDILKIHSRKMNLMRGIDLKKIAEKMSGASGAELKAVCTE
          AGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK"
```

```
BASE COUNT      551 a      293 c      414 g      489 t
ORIGIN
```

```
Query Match    62.1%;   Score 779;  DB 8;  Length 1747;
Best Local Similarity   78.8%;  Pred. No. 4.1e-209;
Matches  929;  Conservative   0;  Mismatches  250;  Indels   0;  Gaps   0;
```

```
Qy    76  AAGCAGGGGCGAGGGCCTCCGCCACTACTAATTCTCTCAACATCCAGGAGCATCAGCTCCTT 135
          ||||||||||||| |||||| || || ||| | ||||||||| |||||||||||||||||
Db   998  GCTACAAACCGGATAGATATTCTTGATCACAAGCTCTTCCTTCCGACCTGGTAGCAGG  975

Qy   136  CTTCGCCAAAAGACTTCATTAACCTTCAAGCCGTCTCGAGGCTCTCAGAGAAAACGAACCTCAAATTCT 195
          |||||||||||||| ||||| ||| | |||||||||| |||| |||||| ||||||||||||||
Db   158  AAGCAATGGCGAGGGGTTTGCCCGCAGTAATTACCTTCAACATATTCCAGGATCTCCCAGCTCCAG 217

Qy   196  AGGGTGAGGATGCTGGCGGAAGAATTACACGCTTCTGCAGGAACCCGGCTGCTTCTTATGTCGGT 255
          ||||| | |||| || ||| || ||| || ||||| |||||||| ||||| | ||| |||||
Db   278  AAAGTGAGGAGGGCTGAGGGAGGGAATTACAGTTGCTCCAGGAACGGGGATCGTATGTTGGT 337

Qy   316  TATGTTGTTGACATTGACAAAAATATTGACATTACAAACGATTACTCCATCCACTAGAGTT 375
          || || || |||||| || ||||| ||||||||| |||| || |||||||||||| |||||
Db   398  TATGTAGTTGATGTAGACAAGAACATTGACATAACAAAAATCNCACCTTCTACAAGAGTA 457

Qy   496  GACCAGCAAAATAAAGAAAATAAAAGAGGTCATTGAGCTACCAATATAAAACATCCTGAGCTG 555
          ||||| |||| |||| ||||| | ||||| ||||| || |||||||| ||||||||||||||
Db   578  GACCAACAGATTAAGGAAAATCAAAAGAGGTTATCGAGCTTCCTATCCAAGCATCCAGAACTA 637

Qy   556  TTTGAAAGTCTTGGAATTGCACAACCAAAAGGGTCTCTGCTCTCTATTGGGCGCCACTGGGTACA 615
          |||||||||||||||||||| ||| || || |||| || || || ||| |||||| |||||||
Db   638  TTTGAAAGTCTTGGAATAGCTCNAACCACCAAAGGGTGTCAACCCTGGCACT 697

Qy   616  GGTAAAACATTGTTGGCCAGCAGTGGCCTCAGGCAGCAGTTGCCATCATCGTACATTCATCAGGGGTG 675
          ||||||||||| ||||||| ||||||||| |||| ||||| ||||||||||||| ||||| |||||||
Db   458  GCTCTTCGTAATGATAGCTACGGTGCTTCAGTTCNGATTTGGCTAGCAAGTAGACCGGTTG 517

Qy   676  TCTGCGTTCTGAGTTAGTTCAGAAATACATTGGAGAAGGAGGGTTCTAGAATAGGTTCAGGCAACTT 735
          |||||||||||||||||||| ||||||||||| || || ||| || ||||| |||||| |||||
Db   758  TCAGGTTCTGAATTGGTTCAGAAGTACATAGCCGAAGTAGGAGCAAGGATTGGTTCGTGGTTCGTT 817

Qy   736  TTTGTTATGGGCCAGGGAACATTGCTCCATCAATTATCTTCAATGGATTGAAATTGACAGTATT 795
          ||||||||| ||||||||| || ||||||||||| |||| || || || | ||||||| |||
Db   796  GGATCTGCTCGGATGGGAATCGTCGGAATCTCCTTCGATTATAATTATTGATGGAAATGACAGTATA 855

Qy   818  TTTGTTATGGCAAGAACAATGCTCCTTCGATTATATTAATTTATTGGATGAAAATTGACAGTATA 877
          |||| | || ||||||| || ||| || || |||| || |||| ||||||||| ||||||||||
Db   878  GGATCCTGCTCGAAATGGAGTCTGGCACTGGTAATGCTTGATAGTGAAGTCCAGCGAGTATG 937

Qy   856  CTGGAACTTCTCAACCACAGTTGGGATGGGATTTGAAGCTTCAAATAGATAGAATTCAAGGTTTGATG 915
          ||||||||||||||||| |||||||| ||||| | ||| || || |||| |||| ||||||| |||||
Db  1298  GAAACAGGAGGAAGAACATGTCACTGAGGAAAACTATGGGAG 1336
```

```
RESULT 13
AY011124
LOCUS         AY011124               1553 bp    mRNA    linear   PLN 28-DEC-2000
DEFINITION    Dactylis glomerata 26S proteasome RPT6a subunit (DGRPT6a) mRNA,
              complete cds.
ACCESSION     AY011124
VERSION       AY011124.1  GI:11993906
KEYWORDS      .
SOURCE        Dactylis glomerata (orchard grass)
  ORGANISM    Dactylis glomerata
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
              Pooideae; Poeae; Dactylis.
REFERENCE     1  (bases 1 to 1553)
  AUTHORS     Alexandrova,K.S. and Conger,B.V.
  TITLE       Identification of a cDNA clone for orchardgrass homolog of the 26S
              proteasome RPT6a subunit
  JOURNAL     Unpublished
REFERENCE     2  (bases 1 to 1553)
  AUTHORS     Alexandrova,K.S. and Conger,B.V.
  TITLE       Direct Submission
  JOURNAL     Submitted (02-NOV-2000) Plant and soil Sciences, University of
              Tennessee, Knoxville, TN 37901-1071, USA
FEATURES             Location/Qualifiers
     source          1..1553
                     /organism="Dactylis glomerata"
                     /mol_type="mRNA"
                     /db_xref="taxon:4509"
                     /tissue_type="embryogenic leaf"
                     /note="authority: Dactylis glomerata L."
     gene            1..1553
                     /gene="DGRPT6a"
     CDS             23..1381
                     /gene="DGRPT6a"
                     /codon_start=1
                     /product="26S proteasome RPT6a subunit"
                     /protein_id="AAG42150.1"
                     /db_xref="GI:11993907"
                     /translation="MAQEEKKPDESTSSRRPQAKETKPNPAPEPEPEMATVAMDISK
                     PPPAAAGDDSAKGARAGGESLRQYLQHIHDLQLQIRTKTHNLRRLEAQRNDLNSKVR
                     ALREELQLLQEPGSYVGEVVKVMGKSKVLVKHPEGKYIVVDIDKSIDITKITPSTRVA
                     LRNDSTMLHLLLPSKVDPLVNLMKVEKVPGSTYDMIGGLDDQQIKEIKEVELPIKHPE
                     LFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKYIGEGSRMV
                     RELFVMAREHAPSIIFMDEIDSIGSARMESGTGNGDSEVORTMLELLNQLDGFEASNK
                     IKVLMATNRIDILDLDQALLRPGRIDRKIEFPNPNEDSRGDILKIHSTTMNLMRGIDLKK"
```

```
                IAGKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAVMKKDNEKMSLRKL
                WK"

BASE COUNT      446 a    349 c    406 g    352 t
ORIGIN

Query Match             61.9%;  Score 775.6;  DB 8;  Length 1553;
Best Local Similarity   76.2%;  Pred. No. 3.7e-208;
Matches  955;  Conservative   0;  Mismatches 299;  Indels  0;  Gaps  0;

Qy      1  ATGGCTCTTGTAGGAGTTGAACTAGAAGAGAGAAGTTGC                                                    60
           ||||| ||  ||  | |  |  |  ||||| |  | ||||
Db    125  ATGGCGACGGGTGGCGATGGCACTCCAAGCCCGGAATTGC                                                   184

Qy     61  TCCGCCAAGCCCACCAAGCAGGGAGGGCGGGGCGCTCCAC                                                   120
Db    185  GCCAAGGGCGCGCCGCGGCCGGGGGGCGGGGCAGCCATCAC                                                  244

Qy    121  GAGCATCAGCCTCCTTCTTCGCCAAAAGACTCATCAGAGA                                                   180
Db    245  GACCTGCAGCTCCAGATCCGGACCAAGACCCACAGGCCCAGGC                                                304

Qy    181  AACGACCTCAATTCTAGGGTGGAGGATGCCTGCCGCGAAGACCC                                                240
Db    305  AACGACCTCAACTCCAAAGTTAGAGACTCAGTTGCTTCAAGAACCT                                              364

Qy    241  GGCTCTTATGTCGGTGAAGTTGTCAAAGTAATGGCCAAGAAGGTC                                               300
Db    305  GGCTCTTATGTTGGTGAGGTGGTGAGGGCTAAATCAAAGGTT                                                  424

Qy    301  CACCAGAAGAGAAAATAGTGTTGACATTGACATTACAGAGATTACT                                              360
Db    425  CATCCAGAGAGGCCAAATAGTCCGATGATATCGACAAGAGCATCGA                                              484

Qy    361  CCATCCACTAGAGTTGTCCACTCCGCAACGACAGTTATGTTCTTCACTTAGTTCTGCCAAGT                               420
Db    485  CCTTCAACAAGAGTTGCTCTTCGAAATGATAGCTACATGCTCCATTTGTCCAAGC                                     544

Qy    421  AAAGTTGATCCATTGGTCAATCTGATGAAAGTTGAGAAAGTTCCCGATTCTACATATGAC                                480
Db    545  AAAGTTGATCCATTGGTCAATCTTAATGAAAGTTGAGAGAAGTTCCAGATTCTACGTATGAT                               604

Qy    481  ATGATTGGTGGTTTAGACAAGCAGCAAATTAAAGAAAATAAAAAGAGGTCATTGAGCCTACCAATC                           540
Db    605  ATGATTGGTGGGCCCTTGATCAGGCAGATCAAAGAGAGTCAAAGCTTATTGAGCTTCCTATC                               664

Qy    541  AAACATCCTGAGCTGTTTGAAGCTCTTGGAATTGCACACCAAAGGGTGCCCTGGTCTCATC                                600
Db    665  AAACATCCTGAGTTATTTGAAGCCCTTGGAATTGCCCAGCCAAAGGGTGTTCTCCTTTAT                                 724

Qy    601  GGGCGCACCTGGTACAGGTAAAAACAATTGTTGGCTCAGGCCAGTGGCTCATCATACTGACTGT                             660
Db    725  GGACCACCAGGGACAGGAAAAACATTATTCGCACCTGCAGTTGCTCACCACTGACTGT                                   784

Qy    661  ACAATTCATCAGGGGTGTCTGGTTCTGAGTTAGTAGTTCAGAAGGTTCTAGA                                         720
Db    785  ACCTTCATCAGAGTATCTGGCTCTGAGTTGGTTCAAAGTATATATTGGTAGGGTTCCCGC                                 844

Qy    721  ATGGTCAGGGAACTTTTTGTTATGGCCAGGGAACATGCTCATCAATTATCTTCATGGAT                                  780
Db    845  ATGGTTCGTGAACTCTTTGTTATGGCTAGGGAGCATGCCACCATCCATTATATTCGATGAT                                904

Qy    781  GAAATTGACAGTATTGGAACCTGCTGCTCGGGATGGCAAGCTGGATGGAGTGATTGAGTGAG                               840
Db    905  GAAATAGACTCCATTGGGATCTGCTAGGGGATTGGAACTGGCACCGGTGATAGTGAA                                    964

Qy    841  GTACAGCGTACTATGCTGGAACTTCCAACCAGTTGGGATTGGAAGCTTCAAAATAAG                                    900
Db    965  GTGCAAGAACCACATGCTTGAGCGTTCTTAACCACTGATGAGGTTTTGAAGCTTCAAACAA                                1024

Qy    901  ATCAAGGTTTTGATGGCCACCAATCGGATTGATATCCTGGATCAAGCCCCTCCTTAGACCA                                960
Db   1025  ATTAAGGTTCTGATGGCAACAAACCGGATAGACATTCTGGATCAAGCCCCTTTGAGGGCCT                                1084

Qy    961  GGACGGGATAGACCCGGAAAATTGAAATTTCCAACCCCTAATGAACGAGAGTCTCGGCGGTTGGATATT                        1020
Db   1085  GGGCCGCATAGACAGGAAGATTGAATTTCCGGAATCCTAAATGAGGATTCACGAGGTGATATC                              1144

Qy   1021  TTGAAAATCCATTCTAGAAGAATGAATTTAATGCGGTGGGCATTGAATTGATTTGAAGAAGATTGCC                          1080
Db   1145  TTGAAGATCCATTCAACAACAATGAACTTGATGCCGTGGAAGTTGAATCCTTAAAAAAAGATCGCA                           1204

Qy   1081  GAGAAGATGAATGGGAGCAATCCTGTGCTTGCTGAACTTAAGGCTCGTGTTGGTTGCACTGAAGCTGAATG                      1140
Db   1205  GGAAAGATGAATGCGTGCTTCAGGAGGTGGTACACGTGACTCAGGAAGATTTTGAGATGGCCGTGCG                          1264

Qy   1141  TTTGCTTTGAGGGGAGGGAGGGGAGGGGTACACGGTGAACTTGATACTACTGACTGAACCTTGAACGTG                        1200
Db   1265  TTTGCTCTCCGTGAGGAGGGTGCATGTAACGCACGAGGAGGACTTCGAAATGGCCGTGGCC                                1324

Qy   1201  AAGGTTGATGAAAAAGGAGACTGAAAAAACATGTCAATTGCGGAAAGTTGTGGAG                                      1254
Db   1325  AAGGTGATGAAGAAGGATAACGAGAAGAACATGTCCTTGCGCAAACTGTGGAAG                                        1378
```

RESULT 14

```
LOCUS       RAT29                   1287 bp    mRNA    linear   ROD 06-FEB-1999
DEFINITION  Rat liver mRNA for proteasomal ATPase (SUG1), complete cds.
ACCESSION   D83521
VERSION     D83521.1  GI:1395176
KEYWORDS    proteasomal ATPase; SUG1.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1287)
  AUTHORS   Makino,Y., Yogosawa,S., Kanemaki,M., Yoshida,T., Yanano,K.,
            Kishimoto,T., Moncollin,V., Egly,J.M., Muramatsu,M. and Tamura,T.
  TITLE     Structures of the rat proteasomal ATPases: determination of highly
            conserved structural motifs and rules for their spacing
  JOURNAL   Biochem. Biophys. Res. Commun. 220 (3), 1049-1054 (1996)
   MEDLINE  96183075
   PUBMED   8607789
REFERENCE   2  (bases 1 to 1287)
  AUTHORS   Makino,Y.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1287)
  AUTHORS   Makino,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-1996) Yasutaka Makino, Chiba University,
            Department of Biology, Faculty of Science; 1-33 Yayoi-cho,
            Inage-ku, Chiba, Chiba 263-8522, Japan
            (E-mail:ymakino@nature.s.chiba-u.ac.jp, Tel:81-43-290-2825,
            Fax:81-43-290-2824)
FEATURES             Location/Qualifiers
     source          1..1287
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /tissue_type="liver"
                     /clone_lib="lambda gt11"
     CDS             34..1254
                     /note="homolog"
                     /codon_start=1
                     /evidence=experimental
                     /product="proteasomal ATPase (rat SUG1)"
                     /protein_id="BAA11938.1"
                     /db_xref="GI:1395177"
                     /translation="MALDGPEQMELEEGKAGSGLRQYYLSKIEELQLIVNDKSQNLRR
                     LQAQRNELNAKVRLLREELQLLQEQGSYVGEVVRAMDKKKVLVKVHPEGKFVVDVDKN
                     IDINDVTPNCRVALRNDSYLHKILPNKVDPLVSLMVEKVPDSTYEMIGGLDKQIKE
                     IKEVIELPVKHPELFEALGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSE
                     IVQKFIGEGARMVRELFVMAREHAPSIIFMDEIDSIGSSRLEGGSGGDSEVORTMLEL
                     LNQLDGFEATKNIKVIMATNRIDILDSALLRPGRIDRKIEFPPPNEEARLDILKIHSR
                     KMNLTRGINLRKIAELMPGASGAEVKGVCTEAGMYALRERRVHVTQEDFEMAVAKVMQ"
```

```
BASE COUNT     344 a    271 c    388 g    284 t
ORIGIN

                      KDSEKNMSIKKLWK"
Query Match
Best Local Similarity   72.4%;  Score 647.8; DB 10;  Length 1287;
Matches  854;  Conservative    0;  Pred. No. 5.9e-172;  Mismatches  322;  Indels   3;  Gaps    1;
```

RESULT 15

```
LOCUS       AB000491               1288 bp    mRNA    linear   ROD 05-FEB-1999
DEFINITION  Rattus norvegicus mRNA for proteasome p45/SUG, complete cds.
ACCESSION   AB000491
VERSION     AB000491.1  GI:2564002
KEYWORDS    .
SOURCE      proteasome p45/SUG.
  ORGANISM  Rattus norvegicus (Norway rat)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (sites)
  AUTHORS   Akiyama,K., Yokota,K., Kagawa,S., Shimbara,N., DeMartino,G.N.,
            Slaughter,C.A., Noda,C. and Tanaka,K.
  TITLE     cDNA cloning of a new putative ATPase subunit p45 of the human 26S
            proteasome, a homolog of yeast transcriptional factor Sug1p
  JOURNAL   FEBS Lett. 363 (1-2), 151-156 (1995)
  MEDLINE   95246863
   PUBMED   7729537
REFERENCE   2  (sites)
  AUTHORS   Leeb,T., Rettenberger,G., Breech,J., Hameister,H. and Brenig,B.
  TITLE     The porcine gene TBP10 encodes a protein homologous to the human
            tat-binding protein/26S protease subunit family
  JOURNAL   Mamm. Genome 3, 180-185 (1996)
REFERENCE   3  (sites)
  AUTHORS   vom Baur,E., Zechel,C., Heery,D., Heine,M.J., Garnier,J.M.,
            Vivat,V., Le Douarin,B., Gronemeyer,H., Chambon,P. and Losson,R.
  TITLE     Differential ligand-dependent interactions between the AF-2
            activating domain of nuclear receptors and the putative
            transcriptional intermediary factors mSUG1 and TIF1
  JOURNAL   EMBO J. 15 (1), 110-124 (1996)
  MEDLINE   96176766
   PUBMED   8598193
REFERENCE   4  (sites)
  AUTHORS   Kazahari,K., Nomoto,K., Nakazato,S. and Ono,M.
  TITLE     Gene coding for the transcription factor, SUG/proteasome, p45 is
            located nearly 40 kb downstream from the rat growth hormone gene
  JOURNAL   Gene 198 (1-2), 323-327 (1997)
  MEDLINE   98036065
   PUBMED   9370298
REFERENCE   5  (bases 1 to 1288)
  AUTHORS   Ono,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-1997) Masao Ono, Rikkyo University, College of
            Science, Department of Chemistry, Life Science Course;
            Nishi-ikebukuro, 3-34-1, Toshima-ku, Tokyo 171-8501, Japan
            (E-mail:mono@rikyo.ac.jp, Tel:+81-3-3985-2387,
            Fax:+81-3-5992-3434)
FEATURES             Location/Qualifiers
     source          1..1288
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /clone="1.3"
                     /sex="male"
```

The remainder of the page consists of a dense nucleotide sequence alignment (Query "Qy" vs Database "Db") with position numbers, running from positions ~76 to ~1095 on the query side and ~1033 to ~1251 on the database side.

```
                     /cell_type="liver"
                     /dev_stage="adult"
                     /note="This sequence is identical with mouse SUG, human
                     p45 and pig TBP10."
     CDS             16. .1236
                     /codon_start=1
                     /product="proteasome p45/SUG"
                     /protein_id="BAA22933.1"
                     /db_xref="GI:2564003"
                     /translation="MALDGPEQMELEEGKAGSGLRQYYLSKIEELQLIVNDKSQNLRR
                     LQAQRNELNAKVRLLREELQLLQEQGSYVGEVVRAMDKKKVLVKVHPEGKFVVDVDKN
                     IDINDVTPNCRVALRNDSYTLHKILPNKVDPLVSLMMVEKVPDSTYEMIGGLDKQIKE
                     IKEVIELPVKHPELFEALGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSE
                     LVQKFIGEGARMVRELFVMAREHAPSIIFMDEIDSIGSSRLEGGSGGDSEVQRTMLEL
                     LNQLDGFEATKNIKVIMATNRIDILDSALLRPGRIDRKIEFPPPNEEARLDILKIHSR
                     KMNLTRGINLRKIAELMPGASGAEVKGVCTEAGMYALRERRVHVTQEDFEMAVAKVMQ
                     KDSEKNMSIKKLWK"
     polyA_signal    1264. .1269
BASE COUNT      347 a    270 c    385 g    286 t
ORIGIN

  Query Match             51.7%;  Score 647.8;  DB 10;  Length 1288;
  Best Local Similarity   72.4%;  Pred. No. 5.9e-172;
  Matches  854;  Conservative    0;  Mismatches  322;  Indels    3;  Gaps    1;

Qy           76 AAGCAGGGCGAGGGCCTCCGCCACTACTATTCTCTCAACATCCACGAGCATCAGCTCCTT 135
                |||   |||   || ||||| || || |||     ||| ||  | || |  ||| |  ||
Db           58 AAGGCAGGCAGTGGACTCCGTCAATATTATCTGTCCAAGATTGAAGAACTCCAGTTGATT 117

Qy          136 CTTCGCCAAAAGACTCATAACCTCAACCGTCTCGAGGCTCAGAGAAACGACCTCAATTCT 195
                 |     | |||| ||| ||||||    | ||  |||| ||||| || || |||||| | 
Db          118 GTGAATGACAAGAGTCAGAACCTCCGGAGACTGCAGGCCCAGAGGAATGAGCTCAATGCA 177

Qy          196 AGGGTGAGGATGCTGCGCGAAGAATTACAGCTTCTGCAGGAACCCGGCTCTTATGTCGGT 255
                |  ||  |  || |||| || ||| | ||||| |||||||||||  ||||| || || || 
Db          178 AAAGTTCGTCTGTTGCGGGAGGAACTGCAGCTGCTGCAGGAACAGGGCTCCTACGTTGGA 237

Qy          256 GAAGTTGTCAAAGTAATGGGCAAGAACAAAGTCCTTGTCAAGGTCCACCCAGAAGGAAAA 315
                ||||| || |  | |  ||  ||||| |||||  | |||||||||| || || || |||
Db          238 GAAGTCGTGAGGGCCATGGATAAGAAAAAAGTGTTGGTCAAGGTCCATCCTGAGGGCAAA 297

Qy          316 TATGTTGTTGACATTGACAAAAATATTGACATTACAAAGATTACTCCATCCACTAGAGTT 375
                | |||||||||  | ||||| || ||||| || |   |  | || ||      | | || 
Db          298 TTTGTTGTTGATGTGGACAAGAACATTGATATCAATGATGTGACACCCAATTGTCGGGTG 357

Qy          376 GCACTCCGCAACGACAGTTATGTTCTTCACTTAGTTCTGCCAAGTAAAGTTGATCCATTG 435
                || ||  | || ||||| ||   ||| ||     | |||| |  ||||| || || |||
Db          358 GCTCTAAGAAATGACAGCTACACTCTGCATAAGATCCTGCCTAACAAAGTGGACCCTTTG 417

Qy          436 GTCAATCTGATGAAAGTTGAGAAAGTTCCCGATTCTACATATGACATGATTGGTGGTTTA 495
                ||    || ||||  || || || || || || || || || || |||||||| ||  ||
Db          418 GTATCACTAATGATGGTGGAAAAGGTGCCAGACTCAACCTACGAGATGATTGGCGGGCTA 477

Qy          496 GACCAGCAAATTAAAGAAATAAAAGAGGTCATTGAGCTACCAATCAAACATCCTGAGCTG 555
                ||| |||| || || || || ||||| || |||||||| ||  | || || |||||||| 
Db          478 GACAAGCAGATCAAGGAGATCAAAGAAGTGATTGAGCTGCCTGTGAAGCACCCTGAGCTC 537

Qy          556 TTTGAAAGTCTTGGAATTGCACAACCAAAGGGTGTCCTGCTCTATGGGCCACCTGGTACA 615
                ||||||   || || |||||||| |||||||| ||||||||| ||||| || || || || 
Db          538 TTTGAAGCCCTGGGCATTGCACAGCCAAAGGGAGTCCTGCTATATGGACCCCCGGGCACT 597

Qy          616 GGTAAAACATTGTTGGCTAGGGCAGTGGCTCATCATACTGACTGTACATTCATCAGGGTG 675
                || || || ||||||||  |||| |||||||||||||| |||||||| || ||  | || 
Db          598 GGGAAGACGTTGTTGGCCCGGGCTGTGGCTCATCATACGGACTGTACCTTTATTCGTGTC 657

Qy          676 TCTGGTTCTGAGTTAGTTCAGAAATACATTGGAGAAGGTTCTAGAATGGTCAGGGAACTT 735
                ||||| |||||  | || ||||||| |||||| |||||  | |||||||| ||||| || 
Db          658 TCTGGCTCTGAACTGGTACAGAAATTCATTGGGGAAGGGGCAAGAATGGTGAGGGAGCTG 717

Qy          736 TTTGTTATGGCCAGGGAACATGCTCCATCAATTATCTTCATGGATGAAATTGACAGTATT 795
                ||||| |||||| | |||||||||||| || || ||||||||||| || || |||  ||| 
Db          718 TTTGTCATGGCCCGAGAACATGCTCCCTCCATCATCTTCATGGACGAGATCGACTCTATC 777
```

```
Qy          796 GGATCTGCTCGGATGGAATCTGGAAGTGGCAACGGTGATAGTGAGGTACAGCGTACTATG 855
                || ||| | ||| ||||    ||   |||    || || |||||||||||||| || |||
Db          778 GGTTCTTCACGGCTGGAGGGAGGCTCTGG---AGGCGACAGTGAGGTACAGCGCACAATG 834

Qy          856 CTGGAACTTCTCAACCAGTTGGATGGATTTGAAGCTTCAAATAAGATCAAGGTTTTGATG 915
                |||||||| ||||||||| |||||| ||||||| ||  | || || ||||||||| | |||
Db          835 CTGGAACTGCTCAACCAGCTGGATGGCTTTGAGGCCACTAAGAATATCAAGGTTATCATG 894

Qy          916 GCCACCAATCGGATTGATATCCTGGATCAAGCCCTCCTTAGACCAGGACGGATAGACCGG 975
                || || ||| ||||||||||||||||    ||||| ||| | || ||  |||| ||| ||
Db          895 GCTACTAATAGGATTGATATCCTGGACTCTGCCCTGCTTCGTCCTGGGAGGATTGACAGG 954

Qy          976 AAAATTGAATTTCCAACCCCTAATGAAGAGTCTCGGCTGGATATTTTGAAAATCCATTCT 1035
                ||||||||||| ||| | || ||||| ||| | ||||||||| |||||||||||||| |||
Db          955 AAAATTGAATTCCCACCTCCCAATGAGGAGGCCCGGCTGGACATTTTGAAAATCCACTCT 1014

Qy         1036 AGAAGAATGAATTTAATGCGTGGCATTGATTTGAAGAAGATTGCCGAGAAGATGAATGGA 1095
                 | | |||||| || |   | || ||  |  ||| ||| ||||| |||  ||||   |||
Db         1015 CGGAAAATGAACTTGACCAGGGGGATCAACCTGAGGAAAATTGCTGAGTTGATGCCAGGA 1074

Qy         1096 GCATCTGGTGCTGAACTTAAGGCTGTTTGCACTGAAGCTGGAATGTTTGCTTTGAGGGAG 1155
                || || || |||||| | ||||  || ||||| || || || | |||| ||||| || |||| 
Db         1075 GCTTCAGGGGCTGAAGTGAAGGGCGTGTGCACAGAGGCCGGCATGTATGCTCTGCGGGAA 1134

Qy         1156 CGGAGGGTACACGTGACTCAGGAGGATTTTGAGATGGCCGTGGCGAAGGTGATGAAAAAG 1215
                ||| | || ||||| ||||||||||| |||||||||||| || || ||||| ||| |||||
Db         1135 CGGCGTGTCCACGTTACTCAGGAGGACTTTGAGATGGCAGTAGCCAAGGTCATGCAAAAG 1194

Qy         1216 GAGACTGAAAAAAACATGTCATTGCGGAAGTTGTGGAAG 1254
                || | ||| ||||||||||  |   |||| | ||||||
Db         1195 GATAGTGAGAAAAACATGTCCATCAAGAAGCTTTGGAAG 1233

Search completed: November  9, 2003, 10:24:27
Job time : 4844 secs
```

```
QM nucleic - nucleic search, using sw model
                    GenCore version 5.1.6
                    Copyright (c) 1993 - 2003 Compugen Ltd.

Run on:        November  9, 2003, 08:46:10 ; Search time 441 Seconds
                                            (without alignments)
                                            7675.953 Million cell updates/sec

Title:         US-09-462-972-1
Perfect score: 1254
Sequence:      1 atggcctcctgctaggagtcga..........cattgcggaagttgtggaag 1254

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0
Searched:      2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters:        5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     N_Geneseq_19Jun03:*
               1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
               2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
               3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
               4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
               5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
               6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
               7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
               8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
               9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
              10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
              11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
              12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
              13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
              14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
              15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
              16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
              17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
              18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
              19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
              20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
              21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
              22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
              23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
              24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
              25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1254	100.0	1254	20	AAX07039	Soybean SUG1 clone
2	826.4	65.9	1476	21	AAC51949	Arabidopsis thalia
3	810	64.6	1476	21	AAC40239	Arabidopsis thalia
4	802.8	64.0	1358	20	AAX07041	Corn SUG1 contig.
5	800	63.8	1477	21	AAC41886	Arabidopsis thalia
6	788.4	62.9	1148	20	AAX07040	Corn SUG1 clone cs
7	633.4	50.5	1505	21	AAC77748	Human cancer assoc
8	633.4	50.5	1800	24	AEO55017	Human ovarian anti

ALIGNMENTS

RESULT 1
AAX07039

```
ID   AAX07039 standard; cDNA; 1254 BP.
XX
AC   AAX07039;
XX
DT   10-MAY-1999  (first entry)
XX
DE   Soybean SUG1 clone sel.pk0023.b5 cDNA insert.
XX
KW   SUG1; transcription factor; coactivator; soybean; 26S proteasome;
KW   ss.
XX
OS   Glycine max.
XX
PN   WO9902689-A1.
XX
PD   21-JAN-1999.
XX
PF   07-JUL-1998;  98WO-US13992.
XX
PR   11-JUL-1997;  97US-0893401.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Allen SM,  Odell JT;
XX
DR   WPI; 1999-120890/10.
DR   P-PSDB; AAW97650.
XX
PT   Newly isolated nucleic acid fragment encoding a plant SUG1 protein
```

Descriptions column (right-hand side):

```
  9  625.4  49.9  1287  17  AAT42343  Human 26S proteaso
 10  608.4  48.5  1396  23  ABL02721  Drosophila melanog
 11  576.8  46.0  1318  23  ABL04117  Drosophila melanog
 12  563.8  45.0  3604  23  ABL02720  Drosophila melanog
 13  541.4  43.2  3373  23  ABL04116  Drosophila melanog
 14  536.6  42.8  1206  24  ABZ32038  Candida albicans e
 15  510.4  40.7  1170  25  ABT19199  Aspergillus fumiga
 16  510.4  40.7  1170  25  ABT21019  Aspergillus fumiga
 17  426.8  34.0  4402  23  ABL15564  Drosophila melanog
 18  324.2  25.9  1351  25  ABT20421  Aspergillus fumiga
 19  324.2  25.9  1351  25  ABT18011  Aspergillus fumiga
 20  324.2  25.9  3350  25  ABT18605  Aspergillus fumiga
 21  324.2  25.9  3351  25  ABT19825  Aspergillus fumiga
 22  312   24.9  3951  19  AAV21209  Methanococcus jann
 23  309.4  24.7  1439  21  AAC42906  Arabidopsis thalia
 24  309   24.6  643   21  AAF14207  Arabidopsis thalia
 25  304.6  24.3  1332  24  ABZ13948  Aspergillus oryzae
 26  304.6  24.3  1476  21  AAC47228  Arabidopsis thalia
 27  303   24.1  1478  21  AAC42167  Arabidopsis thalia
 28  294.4  23.5  1797  21  AAC51097  Arabidopsis thalia
 29  294.4  23.5  2479  25  ABZ36182  Human secretory po
 30  293.6  23.4  3207  23  ABL03176  Drosophila melanog
 31  292.8  23.3  1140  21  AAC41684  Arabidopsis thalia
 32  292.6  23.3  1281  21  AAC51096  Arabidopsis thalia
 33  292.6  23.3  1583  21  AAC33682  Arabidopsis thalia
 34  291.2  23.2  1599  17  AAT14641  S4 gene.  Homo sap
 35  291.2  23.2  1599  24  ABK83793  Human cDNA differe
 36  286   22.8  1407  24  ABK92066  DNA encoding novel
 37  285.8  22.8  1335  24  ABZ32098  Candida albicans e
 38  280   22.3  1081  21  AAC41608  Arabidopsis thalia
 39  279.6  22.3  349980  22  AAF86431  Pyrococcus abyss
 40  275.2  21.9  1398  24  ABZ32393  Candida albicans e
 41  275.2  21.9  1891  24  ABZ35268  Human gene express
 42  272.2  21.7  1891  24  ABZ19352  Aspergillus fumiga
 43  271.8  21.7  1320  25  ABT21172  Aspergillus fumiga
 44  270.8  21.6  1200  21  AAC48647  Arabidopsis thalia
 45  270.6  21.6  1167  18  AAV01892  Human 26S proteaso
```

```
PT   homologue - useful for regulating expression of specific genes
PT   normally controlled by SUGI, and for screening for plant SUGI
PT   protein inhibitors
XX
PS   Claim 7; Page 25-26; 52pp; English.
XX
CC   This polynucleotide comprises the insert of cDNA clone sel.pk0023.b5
CC   which encodes soybean SUGI protein (see AAW97650), a homologue of the
CC   mouse SUGI transcriptional coactivator. The plant SUGI protein is
CC   involved in regulation of gene expression and also functions as a
CC   regulatory component of the 26S proteasome and hence is involved in
CC   the regulation of protein turnover. Clone sel.pk0023.b5 was
CC   isolated from a 5-wk soybean B73 cDNA library on the basis of
CC   homology to ESTs encoding homologues of mouse, yeast and Xenopus
CC   SUGI proteins. Soybean, corn, wheat and rice SUGI polynucleotides
CC   (see AAX07039-43) and polypeptides (see AAW97650-54) are provided, as
CC   well as chimeric genes in which the SUGI polynucleotide is in sense
CC   or antisense orientation, and expression results in production of
CC   altered levels of SUGI protein in transformed host cells. SUGI
CC   proteins can be used to modulate expression of specific genes
CC   whose promoters are normally regulated by SUGI or targeted by
CC   transcription factors that normally interact with SUGI. Host cells
CC   can be used to screen for compounds that modulate SUGI activity.
CC   Isolated SUGI nucleic acids can be used in the recombinant
CC   production of SUGI polypeptides and as sources of probes and
CC   primers.
XX
SQ   Sequence 1254 BP; 367 A; 246 C; 321 G; 320 T; 0 other;

Query Match     100.0%;  Score 1254;  DB 20;  Length 1254;
Best Local Similarity   100.0%;
Matches 1254;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          1 ATGGCTCTTGTAGGAGTTGAAACTGAAGCACCAGAGGATGCGGGACGGGGCGGCGAATTGC   60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCTCTTGTAGGAGTTGAAACTGAAGCACCAGAGGATGCGGGACGGGGCGGCGAATTGC   60

Qy         61 TCCGCCAAGCCACCAAGCGAGGGCCTCCGCACTACTATTCCTCAACATCCAC         120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCCCCCAAGCCACCAAGCGAGGGCCTCCGCACTACTATTCCTCAACATCCAC         120

Qy        121 GAGGATCAGCTCCTTCTTCCGCCAAAAGACTCATAACCGTTCTCGGAGGCTCAGAGA      180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAGGATCAGCTCCTTCTTCCGCCAAAAGACTCATAACCGTTCTCGGAGGCTCAGAGA      180

Qy        181 AACGACCCTCAATTCTAGGGTGAGGATGCTGCGCGGAAGGAATTCACAGCTTACAGCCC      240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AACGACCCTCAATTCTAGGGTGAGGATGCTGCGCGGAAGGAATTCACAGCTTACAGCCC      240

Qy        241 GGCTCCTTATGTCGGTGAAGTTGTCAAACGTAATGGGCAAGAACAAAGTCCTTGTCAAGGTC   300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGCTCCTTATGTCGGTGAAGTTGTCAAACGTAATGGGCAAGAACAAAGTCCTTGTCAAGGTC   300

Qy        301 CACCCAGAGGAGGAAAATATGTGTTGACATTGACAAAAATATTGACAATTACCAAAGATTACT   360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CACCCAGAGGAGGAAAATATGTGTTGACATTGACAAAAATATTGACAATTACCAAAGATTACT   360

Qy        361 CCATCCACTCAGTAGAGTTGCACTCCCGCAACCACAGTTATGTTCTTCACTTAGTTCGTCCAAGT   420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CCATCCACTCAGTAGAGTTGCACTCCCGCAACCACAGTTATGTTCTTCACTTAGTTCGTCCAAGT   420

Qy        421 AAAGTTGATCAATTGGGTCAATTCTGATGAGAAAGTTCCCGATTCTACCATAATGAC   480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AAAGTTGATCAATTGGGTCAATTCTGATGAGAAAGTTCCCGATTCTACCATAATGAC   480

Qy        481 ATGATTGGTGGTTTAGGACCAGCAAATTAAAGAAATTAAAAGAGGTCATTGAGCTACCAAATC   540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGATTGGTGGTTTAGGACCAGCAAATTAAAGAAATTAAAAGAGGTCATTGAGCTACCAAATC   540

Qy        541 AAACATCCTGAGCTGTTTGAAAGTCTTGGAAGTCACACAAAGGGTGTCCTGCTCTAT   600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAACATCCTGAGCTGTTTGAAAGTCTTGGAAGTCACACAAAGGGTGTCCTGCTCTAT   600
```

```
Qy        601 GGGCCACCTGGTACAGGGTAAAAACATTGTTGGGCTGGCTGGCTCATCATACTGACTGT   660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGGCCACCTGGTACAGGGTAAAAACATTGTTGGGCTGGCTGGCTCATCATACTGACTGT   660

Qy        661 ACAATTCATCAGGGTGTCCTGGTTCGTGGTTCGAGTTAGTTCAGAAAATACATTACTTCATGGAT   720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACAATTCATCAGGGTGTCCTGGTTCGTGGTTCGAGTTAGTTCAGAAAATACATTACTTCATGGAT   720

Qy        721 ATGGTCAGGGAACATTTTTGTTATGGCCAGGGAACATGCTCATCAATTATCTTCATGGAT   780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGGTCAGGGAACATTTTTGTTATGGCCAGGGAACATGCTCATCAATTATCTTCATGGAT   780

Qy        781 GAAATTGACAGTATTGGATCTGCTCCGGATGGGAAATCTGGAAGTGGCAACGGTGATAGTGAG   840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTGACAGTATTGGATCTGCTCCGGATGGGAAATCTGGAAGTGGCAACGGTGATAGTGAG   840

Qy        841 GTACAGCGTACTATGCTGGAACTTCTCAACCAGTTGGATGGTTCAAATAAG   900
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GTACAGCGTACTATGCTGGAACTTCTCAACCAGTTGGATGGTTCAAATAAG   900

Qy        901 ATCAAGGTTTTGATTGGTGGCCACCAATCGGATTGATATCCTGATCAAGCCCTCCTTAGAGCCA   960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ATCAAGGTTTTGATTGGTGGCCACCAATCGGATTGATATCCTGATCAAGCCCTCCTTAGACCA   960

Qy        961 GGACCGGATAGACCCGGAAAATTGAATTCCAACCCCTAATGAGAAGAGTCTCGGGCTGGATATT   1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGACCGGATAGACCCGGAAAATTGAATTCCAACCCCTAATGAGAAGAGTCTCGGGCTGGATATT   1020

Qy       1021 TTGAAAATCCATTCTAGGAATGAAATTAATGCGTGGCATTGATTTGAAGAAGAATTGCC   1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTGAAAATCCATTCTAGGAATGAAATTAATGCGTGGCATTGATTTGAAGAAGAATTGCC   1080

Qy       1081 GAGAAGAATGGAATGGAGCAGCATCTGGTGTTGCTGAAACTTAAGGCTGGTTTGCACTGAAGCTTGGAATG   1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAGAAGAATGGAATGGAGCAGCATCTGGTGTTGCTGAAACTTAAGGCTGGTTTGCACTGAAGCTTGGAATG   1140

Qy       1141 TTTGCTTTGAGGGGAGCGGGTACACCTGACTCAGGGAGGATTTTGAGATGGCCGTGGCG   1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TTTGCTTTGAGGGGAGCGGGTACACCTGACTCAGGGAGGATTTTGAGATGGCCGTGGCG   1200

Qy       1201 AAGGTGATCAAAAAGAGGAGAGAGCTGAAAAAACATGTCATTGCGGAAGTTGTGGGAAG   1254
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AAGGTGATCAAAAAGAGGAGAGAGCTGAAAAAACATGTCATTGCGGAAGTTGTGGGAAG   1254
```

```
RESULT 2
AAC51949
ID   AAC51949 standard; DNA; 1476 BP.
XX
AC   AAC51949;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 70076.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;     99US-0121825.
PR   05-MAR-1999;     99US-0123180.
PR   09-MAR-1999;     99US-0123548.
PR   23-MAR-1999;     99US-0125788.
PR   25-MAR-1999;     99US-0126264.
PR   29-MAR-1999;     99US-0126785.
```

PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139919.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139992.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145145.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.

PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160960.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161600.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match   65.9%; Score 826.4; DB 21; Length 1476;
Best Local Similarity   80.8%; Pred. No. 3.4e-246;
Matches 962; Conservative   1; Mismatches 227; Indels 0; Gaps 0;

Qy   65 CCAAGCCCACCAAGCACAGGCGCGAGGGCGCCTAGCGCCCACTACTACTTCTCCAACATCCACGAGC 124
         |||| ||| |||||| |||| || ||||| | | ||||||||||||| ||| |||||||||
Db   181 CCGAGCTGCAAAGCAAAGGGGAAGGGCTTAAGCAGTACTATCTCCAGCACACATCCATGAGC 240

Qy   125 ATCAGCTCCTTCTCCGGCCAAAAGACTCATCAACCTCACCGTCTCGAGGCTCAGAGAAACG 184
         || | ||||| || | ||||||||| ||| | || || ||||||||||||| ||| ||
Db   241 TCCAGCGCCCAGCTTCGACAGAAGACTAATAACCTCAATCGCCTGAAGCTCAAAAGGAATG 300

Qy   185 ACCTCAATTCTAGGGGTGAGGATGCTCGCGGAAGAATTACAGCTTCTGCAGGAACCCGGGCT 244
Db   301 AACTCAATTCTCGAGTGAGGAATGCTCAGAGAAGAAGAGTTACAGCTCCTTCCTTCAAGAAACCTGGGT 360

RESULT 3
AAC40239
ID AAC40239 standard; DNA; 1475 BP.
XX
AC AAC40239;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 27580.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0111825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.

```
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132663.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   17-JUN-1999;   99US-0139453.
PR   18-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139817.
PR   21-JUN-1999;   99US-0139750.
PR   21-JUN-1999;   99US-0139763.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.

PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0147303.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147416.
PR   06-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   09-AUG-1999;   99US-0148171.
PR   10-AUG-1999;   99US-0148319.
PR   11-AUG-1999;   99US-0148341.
PR   12-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   13-AUG-1999;   99US-0149368.
PR   16-AUG-1999;   99US-0149175.
PR   17-AUG-1999;   99US-0149426.
PR   18-AUG-1999;   99US-0149722.
PR   18-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   20-AUG-1999;   99US-0149902.
PR   20-AUG-1999;   99US-0149930.
PR   23-AUG-1999;   99US-0150566.
PR   23-AUG-1999;   99US-0150884.
PR   25-AUG-1999;   99US-0151066.
PR   26-AUG-1999;   99US-0153065.
PR   27-AUG-1999;   99US-0151080.
PR   27-AUG-1999;   99US-0151303.
PR   27-AUG-1999;   99US-0151438.
PR   30-AUG-1999;   99US-0151930.
PR   31-AUG-1999;   99US-0152363.
PR   01-SEP-1999;   99US-0153070.
PR   07-SEP-1999;   99US-0153758.
PR   07-SEP-1999;   99US-0154018.
PR   10-SEP-1999;   99US-0154039.
PR   13-SEP-1999;   99US-0154779.
PR   15-SEP-1999;   99US-0155139.
PR   16-SEP-1999;   99US-0155486.
PR   20-SEP-1999;   99US-0155596.
PR   22-SEP-1999;   99US-0155659.
PR   23-SEP-1999;   99US-0156458.
PR   24-SEP-1999;   99US-0156596.
PR   28-SEP-1999;   99US-0157117.
PR   29-SEP-1999;   99US-0157753.
PR   04-OCT-1999;   99US-0157865.
PR   05-OCT-1999;   99US-0158029.
PR   06-OCT-1999;   99US-0158232.
PR   07-OCT-1999;   99US-0158293.
PR   08-OCT-1999;   99US-0159293.
PR   12-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   13-OCT-1999;   99US-0159329.
PR   13-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159329.
```

PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

Query Match   64.6%;   Score 810;   DB 21;   Length 1475;
Best Local Similarity   80.2%;   Pred. No. 4.4e-241;
Matches 951;   Conservative   0;   Mismatches 235;   Indels   0;   Gaps   0;

Qy   69 GCCCACCAAGCAGGGGCGGAGGGCCGAGGGAAACAATGCTCCATCAATTATCTTCATGGATGAAAATTGA 128
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   146 GGGCGGCGGAAACAAAGGAGGAAGGGCTTAACAGTACTACCTCCAGCATCTGAATGAGAGTCCA 205

Qy   129 GCTCCTTCTTCGGCCAAAAGACTCATAACCTCACCGTCTCGGAGGCTCAGAGAGAAACGGACCT 188
           ||| |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   206 GCGCCTTGCAACGGCGAAAAGTCGTATAACCTTAATCGTCTTGAAGCTCAGAGGAATGAACT 265

Qy   189 CAATTCTTAGGGTGGAGGAATGCTGCGGAGGAATTACAGACCCGGGCTCTTA 248
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   266 CAATTCTTCGGAGTACGAAATGCTTCAGACAAGAGTTACAGCTCCTTCAAGAACCTGGGTCCTA 325

Qy   249 TGTCCGGTGAAGTGTCAAAGTAATGGGCCAAGAACAAAGTCCTTGTCAAGGTCACCCCAGA 308
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   326 TGTGGGTGAAGTGGTAAAAGTGATGGAAAACAAAAATATTGCATTTGGTTAAGGTTCATCCAGA 385

Qy   309 AGGAAAATATGTTTGTTGACATTGACAAAAAATATTGCACTTACAAAGATTACTCCATCCAC 368
           ||||||||||||| |||||||||||| ||||||||||||||||||||||||||||||||||
Db   386 GGGGAAGTATGTTGTCGATAATTGACAAAAAGTATAGCAAACTTACACTCCATCACAAC 445

Qy   369 TAGAGTTGCACTCCCGCAACGACAGTTATGTTCTTCACTTAGTTCTTGCCAAGTAAAGTTGA 428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   446 GAGAGTTGCCTCTCCGTAATGTAAGCTATGTTCTCGCCACTGGTTCTCGCCAAGTAAAGTAGA 505

Qy   429 TCCAATGGTCAATCTGATGAAAGTTGAGAAAGTTCCCGATTCTACATATGACAGCATCC 488
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   506 TCCCTGGTTAACCTTATGAAAGTTGAGAAAGGTTCCCAGACTCCACATATGACATGATTGG 565

Qy   489 TGGTTTAGACCAGCACAAATTAAAGAAATAAAGAGAGGTCATTGAGCTACCAAATCAAACATCC 548
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   566 TGGTTCTTGACCAGCAAAATCAAGGAAATAAAGGAGGTCATTGAACTGCCAATCAAGCATCC 625

Qy   549 TGAGCTGTTTGAAAGTCTTGGAAGTTGCACAACAAAAGGGTGTCCTGCTCTATGGGCCACC 608
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   626 TGAATTGTTTGAGTCTTCTTGGAATTGCGCAACAAGGGGTGTGTTGTTGTACGGTCCACC 685

Qy   609 TGGTACAGGTAAAACATTGTTGGGTAGGGCAGTGGCTCATCATACTGACGTACATTCAT 668
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   686 TGGAACTGGGACAAGACACTATTGGGCTGGCTGGGCACATCACACTGACTGCTTTCAT 745

Qy   669 CAGGGTGTGCTGGTTCTGAGTTAGTTCAGAAATAACATTGGAGAAGAGGTTCTAGAAGATGGTCAG 728
           || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   746 CAGAGTTTCTGGTTCTGAGCTGGTCCCAAGAAATACATTGGAGAAATGGGTTCTAGAATGGTCAG 805

Qy   729 GGAACTTTTTGTTATGGCCAGGGAAACAATGCTCCATCAATTATCTTCATGGATGAAAATTGA 788
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   806 AGAACTTTTTGTTGGATGGCAAGGGAGGAGACGCACCATCAATCATCTTCATGGATGAGAAATGA 865

Qy   789 CAGTAATTGGATCTGCTCGGGATGGAATCTGGAAGTGGCAAGCGTGATAGTGAGGTACACG 848
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   866 TAGTATCGGGTCTGCCTCGTATTGGAAATCTGGTGAGTTGGAAATGGTGACAGTGCAAAG 925

Qy   849 TACTATGCTGGGAACTTCTCAACCACCAGTAGTGGATGGATTTGAAGCTTCAAATAAGATCAGGT 908
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   926 GACTATGCTTGAGCTTCTCAATCAACTTGACGGATTCGAGGGGGTCAAACAAAAATCAAGGT 985

Qy   909 TTTGATGTGGCCACCAAATCGGATTGATTATCCTCTGGATCAAGCTCCTCTTAGACCAGGACCGGAT 968
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   986 TTTGATGGCTACAAAATCGTATTGATTATCTGGATTCAAGTCTTCTCCGGATGGGGTGGAAGGAT 1045

Qy   969 AGACCCGGAAAAATTGAAATTTCCAACCCCTAATGAAGAGTCTCGGCTGGATATTTTGAAAAT 1028
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1046 TGATAGGAAAAAATTTGAATTTTCCTAAATCCATAAATGAAGAGTCACGTTTTGAATATCTTGAAGAT 1105

Qy   1029 CCATTCCTCAGAGAATGAAATTTAAATGCCGTGGGCAGTTGATTTGAAGAAGATTGCCGAGAGAT 1088
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1106 ACACTCAAGGAAAAATGAAATTTGAATGGGTGGGGAATCGATCTGAAAAAGGATTTGCAAGAAGAT 1165

Qy   1089 GAATGGAGCACTCTGGTTGCTGAACTTAAGGCTGGTTGTTTGCACTGAAGTCTGGAATTGTTTTGCTTT 1148
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1166 GAATGGTGCTTCAGGTGCCTCAGGCTGAGCTGAAGCTGTATCGCGAGGCGGCGCATGTTTGCACT 1225

Qy   1149 GAGGGAGGCGAGGGGGTACACCGTGACTCAGCGGAGGAATTTGAGGATGGCCGTGGGGAGGTGAT 1208
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1226 TCGTGAGGAGGAGTACATTGACTCAGGAAGAACTTTGAGCATGGCTGTGGGGCGAAGGTAAT 1285

Qy   1209 GAAAAAGGAGGACTGAAAAAAACATGTCAATTGCGGAAGTTGTTGGAAG 1254
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1286 GAAGAAAGACACAGAGAAGAACAAGATGTCTCTGTAAGCTCTGGGAAG 1331

RESULT 4
AAX07041

ID   AAX07041 standard; cDNA; 1358 BP.
XX
AC   AAX07041;
XX
DT   10-MAY-1999 (first entry)
XX
DE   Corn SUG1 contig.
XX
KW   SUG1; transcription factor; coactivator; 26S proteasome; corn;
KW   maize; ss.
XX
OS   Zea mays.
XX
FH   Key   Location/Qualifiers
FT   CDS   93..1358
FT   /tag= a
XX
PN   WO902689-A1.
XX
PD   21-JAN-1999.
XX
PF   07-JUL-1998; 98NO-US13992.
XX
PR   11-JUL-1997; 97US-0893401.
XX
PA   (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI   Allen SM, Odell JT;
XX
PT   Newly isolated nucleic acid fragment encoding a plant SUG1 protein
PT   homologue - useful for regulating expression of specific genes
DR   WPI; 1999-120890/10.
DR   P-PSDB; AAW97552.

```
PT    normally controlled by SUG1, and for screening for plant SUG1
PT    protein inhibitors
XX
PS    Claim 7; Page 31-32; 52pp; English.
XX
CC    This polynuclectide comprises a contig assembled from the cDNA
CC    inserts in clones csi.pk0051.b7 (see AAX07040), crin.pk0096.b2 and
CC    ctaln.pk0056.c11 isolated from a corn cDNA library on the basis of
CC    homology to mouse, yeast and Xenopus SUG1 ESTs. The encoded SUG1
CC    protein (see AAW97652) is involved in regulation of gene expression
CC    and also functions as a regulatory component of the 26S proteasome
CC    and hence is involved in the regulation of protein turnover.
CC    Soybean, corn, wheat and rice SUG1 polynucleotides (see AAX07039-43)
CC    and polypeptides (see AAW97650-54) are provided, as well as chimeric
CC    genes in which the SUG1 polynucleotide is in sense or antisense
CC    orientation, and expression results in production of altered levels
CC    of SUG1 protein in transformed host cells. SUG1 proteins can be
CC    used to modulate expression of specific genes whose promoters are
CC    normally regulated by SUG1 or targeted by transcription factors
CC    that normally interact with SUG1. Host cells can be used to screen
CC    for compounds that modulate SUG1 activity. Isolated SUG1 nucleic
CC    acids can be used in the recombinant production of SUG1 polypeptides
CC    and as sources of probes and primers.
SQ    Sequence 1358 BP; 376 A; 307 C; 375 G; 300 T; 0 other;

Query Match             64.0%;  Score 802.8; DB 20;  Length 1358;
Best Local Similarity   80.2%;  Pred. No. 7.2e-239;
Matches  942;  Conservative    0;  Mismatches 232;  Indels    0;  Gaps    0;

Qy      81 GGGCGAGGGGGCCTCCGCGACCTACTATTCTCTCAACATCCACGAGCATCAGCTTCCTTCTTCG 140
           ||||| ||||||| ||||||||    |||  ||||||| || ||||||||||||||| || ||
Db     185 GGGCGCAGGGGGCTGCGGCCAGTGACTACCTGCAGCACATCCATGAGCTTCCAGATCCG 244

Qy     141 CCAAAAAGACTCATAACCTCAACCGTCTCGAGGGTCTCGAGGAAAACGACCCTCAAATTCTAGGGT 200
           |||||   |||| ||||||| |||||  |||  |||||||| || ||  ||  ||||| |||||||
Db     245 GCAGAAGACCCATAACCTCAACCGGCCTCGAGGCCCGACCTCAACTCCCGAGT 304

Qy     201 GAGGGATGCTGCCGCGAAGAAATTACAGCTTCTGCCAGGACACCCGGCTCTTATGTCGGTTGAAGT 260
           |||   | |||||  | ||  |||||| |||||||||| |||||| |||| |||||| |||||||
Db     305 TAGAATGCTCAGGGAAGAGAGTTGCACTTGCTTGTTCAACAGAGCCTTCACTATAGTTGGTGAGGT 364

Qy     261 TGTCAAAGTAATGGGCGAAGAACAAAGTCCTTGTCAAGGTCCACCCACGAGAGGAAAATATGT 320
           | |||||| || |||||||||||||||  ||  ||||||||||||   |||||||||||||||
Db     365 GGTGAAGGTCATGGGGGAAATCAAAGGTTCTGGGTGAAGGTACACCCCGGAGGCAAATATGT 424

Qy     321 TGTTGACATTGACAAAAATAATTGACATTGCAAAGAATTACTCCATCCACTAGAGTTGCACT 380
           || ||||||||||| ||||||||||||||  |||||||||| |||||||||||||||||||
Db     425 GGTGGATATAGATAAGACAGCATTGAATATCAACACACCTTCAACAAGAGTTGCTCT 484

Qy     381 CCGCAACGACAGCAGTTATGTTCTTCACTTAGTTGTCGCCAAGTAAAGTTGATCCATTGGTCAA 440
           |||| | | |||| ||| |||| |||||||||| | |||  |||||||| || ||||||||||
Db     485 TCGGAATGACAGCTATATGCTCCATCCTGATCCTACCAAGCAAAGTTGATCCATTTGGTCAA 544

Qy     441 TCTGATGAAAGTTGAGAAAGTTCCCGGATTCTACATATGACATGATTGGTTGGTTTAGACCA 500
           |||| ||||||||||||||||  || |||||  |||||||| || | | ||| || ||||||
Db     545 TCTCATGAAAGTTGAGAAGGTTCCGGAATTCAACCTATGATAATGGGAGGCCCTGACCA 604

Qy     501 GCAAATTAAAGAAAATAAAGAGGTCATTGAAGCTACCAAATCAAACATCCTGAGCTGTGTTTTGA 560
           |||||||| | | |||||| ||||||||||||  ||| |||   || ||  || |||| |||||
Db     605 GCAAATTAAAGAAGATCAAAAGAGGTCATTGAGCTTCCAATCAAACATCCCGGAACTGTTGA 664

Qy     561 AAGTCTTTGGAATTCGCACAACCAAAGGGGTCGTTCCTCTGCGCTTGGTACAGGTAA 620
           |  ||||||||||| ||| || |||  || ||| |||||||  || ||| | |||||
Db     665 GAGCCTTTGGAATTCGCACAACCAAAGGGGGTGTCCTTGGGCACACAGGAAA 724

Qy     621 AACATTGTTGGCTAGGGCAGTGGCTCATCATACTGACTGTACATTCATCAGGGTGTCTGG 680
           || |||||| ||| || |  || |||||||| ||| |||| || || | || |||||||
Db     725 GACATTGTTGGCACGTGCCGGTTGCTCATCACACTGACCCTTCATCAGGGTGTCTGG 784

Qy     681 TTCTGAGTTAGTTCAGAAATAACATTGGACGAGAAGGTTCAGAGAACTTTTTTGT 740
           ||||| |||| || |||||| ||| ||| || ||||||| || ||  ||| |||
Db     785 TTCTGAGTTGGTTCAGAAGATATATTGGTGAGGGCGTTCCCGGATGGTTAGGGAACTCTTTGT 844
```

```
Qy     741 TATGGGCCAGGGAACAATGCTTCCATCCAATTCATTATCTTCATGGATGAAATTGACAGTAGTATTGGATC 800
           ||||||||||||||||||  |||||  ||||||||||||||||||||| |||| |||||  ||||| |||||
Db     845 TATGGGCCAGGGAACCATGCCACCCATCCAATTCATTATCTTCATGGATGAAAATGACTCTCTATCCGGATC 904

Qy     801 TGCTCGGATGGAATCTGGAAGTGGCAACGGTGATAGTAGTGAGGGTACAGGGTACTATGCTGA 860
           ||| |||||| ||||||||||||| ||| ||||| || ||||| || ||| ||| |||||||
Db     905 TGCTAGAATGGAAGTCTGGAACTGGTGAACTGAGCAGTGAAGTTCAGCGGTACTATGCTTGA 964

Qy     861 ACTTCTCAACCAGTTGGATGGAATTTTGAAGCTTCAAATAAGATCAACAAGTTTTGATGGCCAC 920
           ||||| |||||||||| |  |||| |||||||||| || || ||| | ||| |||||||| |
Db     965 ACTTCTAAACCAGCTCGATGGTTTTGAAGCATCAAACAAAAATTAAGGTTTTGATGGCAAC 1024

Qy     921 CAATCCGGATTCATATCCTGGATCAAGCCCTCCTTAGACCAGGACGGATAGACCGGAAAAT 980
           |||||  |||| ||||||||||||| ||  ||| ||||||  ||||||||||| |||||||
Db    1025 GAACAGAATAGACATTTGGATCAACCCCTCCTTCGAGCCCTGGATCCTGGACATAGACAGGAAGAT 1084

Qy     981 TGAAATTCCAACCCCTTAATGAAGAGTCTCGGCGTGCGATATTTTGAAAATCCATTCTAGAAG 1040
           ||||||||| ||  |||||| |||||| | |||  | || |||||||||  ||||||||||||
Db    1085 TGAATTCCAAATCCTAACGAGGATTCACGTTTCGGATATCTGAAGATCCATTCACAGAAA 1144

Qy    1041 AATGAATTTAATGCGTGGCATTGATTTGAAGAAGATTGCCCGAAGAGAAGATGAAGTGGACATC 1100
           ||||| | ||| |||  |||| | | ||||||| |||||| |  | ||| |||| ||| ||||
Db    1145 AATGAACTTGATGCGTTGGCCATTGACTGAAAAGATCGCGGGAAAAGATGAATGGGCCGCTC 1204

Qy    1101 TGGTGCTGAACTTAAGGCTGGTTTTGCACTGAAGCTGGAAGCTGAAATGTTTTGCTTGAAGTTGAGGGAGGCGGAG 1160
           |||| ||||| | ||||||||| ||||  || || |||| |||||  |||| |||||||||||  | ||| |||||
Db    1205 AGGAGGTGAGCTCAAGGCCGTCTGCGACAGAGGCCTGGAAATGTTTGCTCTTGAGGTTGAAG 1264

Qy    1161 GGTACACGTGACTCAGGAGGGANTTTGAGATGGCCGTGGCGGAAGGTGAATGAGAAAAAGGAGAC 1220
           |||| ||||||| ||| ||||  |||| || ||||| |   |||| || | | || || ||||
Db    1265 GGTGCACGTTACCCAGGAGGACTTCGAGATGGCAGTGGCCAAGGTGATGAGAAAAGACAC 1324

Qy    1221 TGAAAAAAACATGTCATTGGGGAAGTGTCTGGAAG 1254
           ||| | | |||||  ||||||||||||| |||||
Db    1325 GGAGAAGAACATGTCCCTGCGCGCAAGCTCTGGAAG 1358
```

Qy   741 ...

RESULT 5

```
AAC41886
ID   AAC41886 standard; DNA; 1477 BP.
XX
AC   AAC41886;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 33500.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000;  2000EP-0301439.
XX
PR   25-FEB-1999;  99US-0121825.
PR   05-MAR-1999;  99US-0123180.
PR   09-MAR-1999;  99US-0123548.
PR   23-MAR-1999;  99US-0125788.
PR   29-MAR-1999;  99US-0126264.
PR   25-MAR-1999;  99US-0126785.
PR   01-APR-1999;  99US-0127462.
PR   06-APR-1999;  99US-0128234.
PR   08-APR-1999;  99US-0128714.
PR   16-APR-1999;  99US-0129845.
PR   19-APR-1999;  99US-0130077.
PR   21-APR-1999;  99US-0130449.
PR   23-APR-1999;  99US-0130510.
```

PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132494.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 18-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139464.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139763.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150586.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 30-AUG-1999; 99US-0151066.
PR 31-AUG-1999; 99US-0151080.
PR 01-SEP-1999; 99US-0151303.
PR 07-SEP-1999; 99US-0151438.
PR 10-SEP-1999; 99US-0151930.
PR 13-SEP-1999; 99US-0152363.
PR 15-SEP-1999; 99US-0153070.
PR 16-SEP-1999; 99US-0153758.
PR 20-SEP-1999; 99US-0154018.
PR 22-SEP-1999; 99US-0154039.
PR 23-SEP-1999; 99US-0154779.
PR 24-SEP-1999; 99US-0155139.
PR 28-SEP-1999; 99US-0155486.
PR 29-SEP-1999; 99US-0155659.
PR 04-OCT-1999; 99US-0156458.
PR 05-OCT-1999; 99US-0156596.
PR 06-OCT-1999; 99US-0157117.
PR 07-OCT-1999; 99US-0157753.
PR 08-OCT-1999; 99US-0157865.
PR 12-OCT-1999; 99US-0158029.
PR 13-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 14-OCT-1999; 99US-0159293.
PR 14-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.

```
PR   14-OCT-1999;  99US-0159638.
PR   18-OCT-1999;  99US-0159584.
PR   21-OCT-1999;  99US-0160741.
PR   21-OCT-1999;  99US-0160767.
PR   21-OCT-1999;  99US-0160768.
PR   21-OCT-1999;  99US-0160770.
PR   21-OCT-1999;  99US-0160814.
PR   21-OCT-1999;  99US-0160815.
PR   22-OCT-1999;  99US-0160980.
PR   22-OCT-1999;  99US-0160981.
PR   22-OCT-1999;  99US-0160989.
PR   25-OCT-1999;  99US-0161404.
PR   25-OCT-1999;  99US-0161405.
PR   25-OCT-1999;  99US-0161406.
PR   26-OCT-1999;  99US-0161359.
PR   26-OCT-1999;  99US-0161360.
PR   26-OCT-1999;  99US-0161361.
PR   26-OCT-1999;  99US-0161920.
PR   28-OCT-1999;  99US-0161992.
PR   28-OCT-1999;  99US-0161993.
PR   29-OCT-1999;  99US-0162142.
```

```
Query Match             63.8%;  Score 800;  DB 21;  Length 1477;
Best Local Similarity   80.7%;  Pred. No. 5.7e-238;
Matches  957;  Conservative    1;  Mismatches  226;  Indels    2;  Gaps    2;
```

```
Qy     69  GCCCACCAAAGACAGAGGCGAGGGCCTCCCCGACTCACTATTCTTCTCAACATCCACGAGCATCA  128
           |||  |  |||| || ||| || ||||  |  | || || | | || ||||| |||| |||| ||
Db     69  GGCGGGCCGAAACAAGGGCGGAGGGCTTAAGCAGTACTAATCCTCAGCACATCCATGAGCTCCA  128

Qy    129  GCTCCTCTTCGCCAAAAGACTCATAAACCTCAACCGGTCTCGAGGCCTCAGAGAAACGACCT    188
           |  ||  || ||| || |||| | || | | |||||| |  |  ||||  |||||| || ||
Db    129  GGGGAAGTATGTTGTTGAATATTGACATTACACAAGTTACTCCATCCAC                 188

Qy    189  CAATTCTTAGGGTGAGGAGTGCTCCGCGAAGAATTACAGCTTCTGCAGGAACCCGGGTCTTTA   248
           |  |||| || | | || || || | |  |||| ||| | |||| ||  |  ||||| ||| |
Db    243  CAATTCTCGAGTGAGAATGCTCAGAAGAGTTACAGCTCCTCCTTCAAGAACTACTGGGGTCCTA  301

Qy    249  TGTCGGTGAAGTTGTCAAAGTAAGTGGGCCAAGAACAAAGTCCTTGTCAAGGTCCACCCAGA    308
           |  ||| ||| ||| | ||| || || ||| || || ||||| | ||| ||||| | |||||
Db    302  GAGAGTTGCTCTTCGTAATGATAGCTATGTTCTCCACCTGGTTCTGCCAAGTAAAGTAGA      361

Qy    309  AGGAAAATATGTTGTTGTTCAACATTGACAAAAAATATTGACATTACACAAGATT CATCCAC   368
           | |  | || | || || | | |  || | | | | | | || | || | | | || |||
Db    362  TGTGGGAGAAGTGGGTAAAAGTTAATGGGCCAAGAAAAACAAAGTCCTTGAAGGTTCATCCAGA  421

Qy    369  TAGAGTTGCACTCCCGCAACGACAGTTATGTTCTTCACTTAGTTCTGCCAAGTAAGTTGA     428
           ||  || || || |||  ||||  | | | | |||||| |||  | |||||  |||||||
Db    422  GGGGAAGTATGTTGTTGAATATTGACATTACACAAGATTCACTCCATCAAC               481

Qy    429  TCCATGGTCAATCTGATGAAAGTTCAGAAAGTTCCGATTCTACATATGACACATGAGATTGG   488
           | |  ||| || | | || |||| | ||| || || | || | | |||| | | || |||
Db    482  GAGAGTTGCTCTTCGTAATGATAGCTATGTTCTCCACCTGGTTCTGCCAAGTAAAGTAGA      541

Qy    489  TGGTTTAGACCAGCAAATTAAAGAAATAAAAGAGGTTCATTGAGCTACCAATCAAAACATCC   548
           ||| || || ||| || || || || |  || | ||| || |||||| || |  || ||||
Db    542  TCCCTTGGTTAACCTTATGAAAAGTTGAGAAGGTTCCAGACTCACCTTATGACATGATTGG     601

Qy    549  TGAGCTGTTTGAAAGTCTTGGAATTGCACAACCAAAGGGTGTTCCTGCTTCATAGGGCCACC   608
           || | | ||| ||| | | ||| ||| | | |||| || || |  | || | ||| |||||
Db    602  TGGTTCTTGACCAGCAGAATTAAGGAGATTAAAGGAGGTTCATTTGAGCGTCAAGCAGCATCC  661

Qy    609  TGGTTACAGGGTAGAAAACATTGTTGGGCTAGGCAGTGGCTCATCATACTGTACACTTCAT    668
           ||| || || | || | || |  | | || || || || | || | | |||| | || |||
Db    662  TGGTTACAGGGTAGAAGTTCACGTGACACACAAGAGACTTGAGATGGCGTGGCCCAAGGTAAT  721

Qy    669  CAGGGTGTCTGGTTCTGAGTTAGTTCAGAAATATCATTGGAGAAGGTTCTGAGAATGGTCAG   728
           ||| | | | ||||| | | | ||||| ||| |  ||| |||| || | || ||||||| ||
Db    722  TGGAACTGGGAAGACACTATTGGCTCGGGCGTCTGTTGCACATCACACTGACATGACTTTCAT  781

Qy    729  GGAACTTTTTTGTTATTGGGCGGAACTACAGCTTCCATCCAATTATATCTTCATCGGATGA    788
           || ||| ||  ||| | ||| |||| | |||  | |||  ||||| ||  || | |||||||
Db    782  AAGACTTTCTGAATTCCGAGCTTGTCCAGGAAATACATTGGAGAGGTTCAGATGAAGTGGTCAG  841

Qy    782  AAGACTTTTCTGAATTCCGAGCTTGTCCAGGAAATACATTGGAGAGGTTCAGATGAAGTGGTCAG  841
           |||| ||| | ||| | || | |||  ||| | | |||| || ||  || | ||||||||||||
Db    842  AGAACTTTTGTGATGCGACAAGGAGCAATGCAACTCAATCTTCAATCTTCATGGATGAAATCGA  901
```

SUG1; transcription factor; coactivator; 26S proteasome; corn;
maize; ss.

RESULT 6

AAX07040

```
ID   AAX07040 standard; cDNA; 1148 BP.
XX
AC   AAX07040;
XX
DT   10-MAY-1999  (first entry)
XX
DE   Corn SUG1 clone csl.pk0051.b7 cDNA insert.
XX
KW   SUG1; transcription factor; coactivator; 26S proteasome; corn;
KW   maize; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FH
FT   CDS             3..1148
FT                   /*tag= a
XX
PN   WO9902689-A1.
XX
PD   21-JAN-1999.
XX
PF   07-JUL-1998;  98WO-US13992.
XX
PR   11-JUL-1997;  97US-0893401.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Allen SM, Odell JT;
XX
DR   WPI; 1999-120890/10.
DR   P-PSDB; AAW97651.
XX
PT   Newly isolated nucleic acid fragment encoding a plant SUG1 protein
PT   homologue - useful for regulating expression of specific genes
PT   normally controlled by SUG1, and for screening for plant SUG1
PT   protein inhibitors
XX
```

```
Qy    789  CAGTATTGGAATCTGCTCCGGATGGAATCTTGGAACTGGCCAACGGTCAGATAGTAGTGAGGTTACACGC   848
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    902  TAGTATTGGGTCGTCCGCTCCTGTATGACAATCTTGGAACTGGAAATGGTGACAGTGAGGTTGCCAACG  961

Qy    849  TACTATGCTGGAACTTCTCAACCAGTTGGATGGATTGAAGCTTCAAATAAGATCAACAAGGT   908
           ||||| |||||||| ||| || ||||| |||| || |||| ||||| | || ||||||| ||
Db    962  GACCAATGCTCGACGCCTTCTCCAATCAACTAACGGGTTTTGAAGCCTCAAACAACAATCAAGGT   1021

Qy    909  TTTGATGCCCACCACCAATCGAGATTGATATATCCTGATCAGCCTCCCTTAGACCAGGACGGAT  968
           |||||| ||||||| | || || | |  | || || || |||||| | || |  || || || |
Db   1022  TTTGAATGGCTACAAATCGTATTCGATTGATATTCTCGGGAATCAAGCTCTTCTCAGCCTGGAAGAAT  1081

Qy    969  AGACCGGAAAAATTGAAATTCCAACCCCTTAATGAAGAAGAGTCTCGGCCTGGATATATTTGAAAAT   1028
           ||  | ||||| || ||||| || | || ||| |||| ||||| || ||  | || | |||| |||
Db   1082  TGAATAGGAAAAATTGAAAATTCCTTAATCCTTAATGAAGAGAGTCAGTTTTGATCTTGGAAGAT  1141

Qy   1029  CCAATTCTAGGAAGAAATGAAATTTAATTGCGGTGCCATTGATTTTGACGAAGTTGTGGAAG   1088
           |||| || | ||||||||||||  | | | ||||| | ||||| || ||  || | || ||
Db   1142  TCACTCGGAGGAAAAATGAAATTTGATGGCGGTAAGCTGAAATGGCAGGAAGAT   1201

Qy   1089  GAATGGAGACATCTCGGTCGTTGCTGACTTAAGGCTGTTTGCACTGAAGCTGGAAATGTTTTGCTTT   1148
           ||||| |||| | || | || |  | || || || || |||| || |||||| | || || ||||
Db   1202  GAATGGCCGCCTTCAGGTGCTCAGGCTCTTCCAGGCTGTTTGCTTCTCAGCCCTGGAAGT   1260

Qy   1209  GAAAAAGGAGGACTGAAAAAAAACATGTCATTGCGGAAGTTGTGGAAG   1254
           |||| |||| || || ||| |   | || | || || || | || |||
Db   1321  GAAGAAGACACGGGAGAAGAACATGTCTCTGCGGTAAGCCTGTGGAAG   1366
```

PS   Claim 7; Page 28-30; 52pp; English.
XX
CC   This polynucleotide comprises the insert of cDNA clone cs1.pk0051.b7
CC   which encodes corn SUGI protein (see AAW97651), a homologue of the
CC   mouse SUGI transcriptional coactivator. The plant SUGI protein is
CC   involved in regulation of gene expression and also functions as a
CC   regulatory component of the 26S proteasome and hence is involved in
CC   the regulation of protein turnover.   Clone cs1.pk0051.b7 was
CC   isolated from a corn cDNA library on the basis of homology to ESTs
CC   encoding homologues of mouse, yeast and Xenopus SUGI proteins.
CC   Soybean, corn, wheat and rice SUGI polynucleotides (see AAX07039-43)
CC   and polypeptides (see AAW97650-54) are provided, as well as chimeric
CC   genes in which the SUGI polynucleotide is in sense or antisense
CC   orientation, and expression results in production of altered levels
CC   of SUGI protein in transformed host cells. SUGI proteins can be
CC   used to modulate the expression of specific genes whose promoters are
CC   normally regulated by SUGI or targeted by transcription factors
CC   that normally interact with SUGI. Host cells can be used to screen
CC   for compounds that modulate SUGI activity. Isolated SUGI nucleic
CC   acids can be used in the recombinant production of SUGI polypeptides
CC   and as sources of probes and primers.
XX
SQ   Sequence 1148 BP; 330 A; 240 C; 302 G; 276 T; 0 other;

Query Match   62.9%; Score 788.4; DB 20; Length 1148;
Best Local Similarity   80.6%; Pred. No. 2e-234;
Matches 921; Conservative   0; Mismatches 221; Indels   0; Gaps   0;

Qy   113 ACATCCACCAGAGCATCAGCTCCTTCCTTCCCAAAAGACTCATAACCTCCAACCGTCTCGAGG 172
            ||||||||||||||| | ||||||||||||||||||||||||||||||||||||||||||||
Db   7 ACATCCATGACCTCGCAGCTCCAGAGATCCGGCAGAGACCCATAACCTCCAACCGTCTCGAGG 66

Qy   173 CTCGAGAAAACGACCTCAATTCTTAGGGTGAGAGAATGCTGCGGCGAAGAAAGTTACAGCTTCTGC 232
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   67 CCCAGCGCAACGACCTCAACTCCCGAGTTAGAATGCTCAGGGAGAGTTGCCAGTTGCTTC 126

Qy   233 AGGAACCCGGCCTCTTATCGTCGGTGAGAGTTGTCAAAGTAATGGGCAAGAACAAAGTCCTTG 292
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   127 AAGAGCCCGCTCATATGTTGGTGAGGTGGTGAAGGTCATGGGGAAATCAAAGGTTCTGG 186

Qy   293 TCAAGGTCCCACCCAGRAGGAAAAATAATGTTGTTGACATTGACAAAAATATTGACATTACAA 352
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   187 TGAGGTACATCCCGAAGGCAAATATGTCGGTGGATAATAAGAGACATTGATATCACTA 246

Qy   353 AGATTACTTCCATCCACTAGAGTTGCACTCCGCAACGACGACAGTTTATGTTCTTCACTTAGTTC 412
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   247 AGATCACACCTTCAACAAGAGAGTTGCTCTTCGGAATGACAGCTTATATGCTCCCATCTGATCC 306

Qy   413 TGCCAAGTGAAAGTTGATCCATTGGTCAATCTGATGAAAGTTGGAGAAGTTCCGATTCTA 472
            | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   307 TACCAAGCAAGTTGATCCATTGGTCAAATCTCAGCAGAAAGTTGGAGAACGTTCCGGATTCTA 366

Qy   473 CATAATGACAATGAATTGGGTGGTTTAGACCAGCAAATTAAAGAAATAAAAGAGGTCATTGAGC 532
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   367 CCTATGATATTGAATTGGGAGCCTTGACCACCAAATTAAAGAAGATCAAAGAGGTCATTGAGC 426

Qy   533 TACCAATCAAACAATCCTGAGCCTGTTTGAAAGTCTTGGAATTGCACAACCAAAGGGTGTCC 592
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   427 TTCCAATCAAACATCCGGAAACTGTTTGAGAGGCCTTGCGCAACCAAAGGGTCC 486

Qy   593 TGCTCTATGGGCCACCTGGTACAGGTAAAACATTGTTGGCTAGGGCAGTGGGCTCATCATA 652
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   487 TTCTTTATGGACTCCGGCCACAGGAAAGACATTGTTGCACTGCCGGTTGCTCATCACA 546

Qy   653 CTGACTGTACATTCATCAGGGTGTCTCGGTTCTCGAGTGTTAGTTCAGAAAATACATTGGAGAAG 712
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   547 CTGACTGCACCTTCCATCAGGGTGTCTCGGTTCTCGAGTTGGTTCAGAAGTATAATTGGTGAGG 606

Qy   713 GTTCTAGAATGGTCAGGGAGAACTTTTTGTATATGGCCAGGACAGGTCACTCATTTATTCT 772
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   607 GCTCCCCGGATGGTTAGGGAGAACTCTTGTTATGGCCAGGACATGCCAACATCCATTATAT 666

Qy   773 TCATGGATGAAAATTGACAGTATTGGATCTGCTCGGAGTGGCCAACGGTG 832
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db   ...

Db   667 TTATGGATGAAAATTGACACTGTCTTCATTCGGATCTGCTAGAATGAGTCTGGAACGGTG 726
            ||||||||||||||||| ||| |||||||||| |||||||||||||| ||||||||||||
Qy   833 ATAGTGAGGTACAGCCGTACTATGCTTGGAACTTCTCAACCAGTTGGATGGATTGGAAGGTT 892
            |||||| |||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   727 ACAGTCAAGTTCAGCGTACTATGCTTGAACTTCTAAACCAGCTCGATGGTTTTGAAGCAT 786

Qy   893 CAAATAAGATCAAAGTTTTGATGCCACCAATCGGATTGATAATATCCTGATCAAGCCCTCC 952
            |||||||||||||||||||||| |||||||||||| ||| |||||||||||||||||||||
Db   787 CAAACAAAATTAAGGTTTTGATGGCAACGAAGATAGACAATTTGGATTCAAGCCCTTC 846

Qy   953 TTAGACCAGACGACATAGACCCGGAAAAAATTTGAATTTCCAACCCCTTAATGAAGAGTCTCGC 1012
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   847 TGAGGCCCTGGCCCGACAGAGACACAGGAAGATTGAATTCCAAAATCCTAACGGAATCACGTTt 906

Qy   1013 TGGATATTTTGAAAATCATTCTTAGAAGAAATGAAATTTAAATGCCGTTGGCATTGATTGAGA 1072
            ||||||||| || || |||||||||||| |||||| |||||||||||||||||||||||||
Db   907 TCGATATCTTCAAGATCCATTCAAGAAAATGAACTTGATGCGTGCCATTGATCTGAAAA 966

Qy   1073 AGATTGCCGAGAAGATGAATGGCAATCTGGGTGCTGAACTTAAGGCTGTTTGCACTGAAG 1132
            |||||| |||||||||||||||||||||||||||||||||||||| ||||| ||||||||
Db   967 AGATCGCCGGAAAAGATGAATGGGCGCCCTCAAGGAGTCGAGCTTCAAGGGCCGTCTCAGCAGAGG 1026

Qy   1133 CTGGAATGGTTTGCTTTGAGGGAGCGGGAGGGTACACGTGACTCAGGAGGAGGATTTTGAGATGG 1192
            ||||||||||||||||||||||||||||| |||| || |||||||||||||||||||||
Db   1027 CTGGAATGTTTGCTTTCGGTGAGGAGCGGAGAGGGTGCAGGTAGCCGGTACCCAGGAGGACTTCGAGATGG 1086

Qy   1193 CCGTGCGCGAAGGTGATGAAAAAGCGAGACTGAAAAAAAACATGTCATTTGCCGGAAGTTGTGGA 1252
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1087 CAGTGGCCCAAGGTGATGAGAAAGAGAGACACGGAGAGAGAACATGTCCCGCGCAAGCTCTGGA 1146

Qy   1253 AG 1254
            ||
Db   1147 AG 1148

RESULT 7
AAC77748
ID   AAC77748 standard; cDNA; 1505 BP.
XX
AC   AAC77748;
XX
DT   08-FEB-2001 (first entry)
XX
DE   Human cancer associated gene sequence SEQ ID NO:142.
XX
KW   Human; cancer associated gene; cancer antigen; detection; cancer;
KW   diagnosis; cytostatic; proliferative; vulnary; immunomodulator;
KW   antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW   antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW   dermatological; neuroprotective; antipsoriatic; thrombolytic; nootropic;
KW   vasotropic; antiangiogenic; gene therapy; inflammation;
KW   immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW   allergic reaction; graft versus host disease; organ rejection;
KW   haemostatic; thrombolytic; cardiovascular disorder; infection;
KW   neurological disease; drug screening; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055350-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05882.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,   Ruben SM;
XX
DR   WPI; 2000-587533/55.
DR   P-PSDB; AAB43539.

```
XX
PT   Novel isolated nucleic acids comprising sequences encoding peptides
PT   useful for treating or diagnosing e.g. cancer.
XX
PS   Claim 1; Page 722; 2352pp; English.
XX
CC   AAC77607 to AAC78448 encode the human cancer associated proteins given
CC   in AAB43398 to AAB44229. The proteins can have activities based on the
CC   tissues and cells the genes are expressed in. Example of activities
CC   include: cytostatic; proliferative; vulnerary; immunomodulator;
CC   antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC   antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC   dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC   nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC   polynucleotides and polypeptides can be used for preventing, treating or
CC   ameliorating medical conditions and diagnosing pathological conditions.
CC   Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC   the present invention may be used to treat immune disorders by activating
CC   or inhibiting the proliferation, differentiation or mobilisation of
CC   immune cells, to treat disorders of haematopoietic cells, autoimmune
CC   disorders, allergic reactions, graft versus host disease and organ
CC   rejection, modulate haemostatic or thrombolytic activity, modulate
CC   inflammation, cancers, cardiovascular disorders, neurological disease and
CC   bacterial or viral infections. The peptides, nucleotides, antibodies,
CC   agonists and antagonists may be also be used in drug screens. AAC78449 to
CC   AAC78457 and AAB44240 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 1505 BP; 434 A; 333 C; 430 G; 304 T; 4 other;

Query Match             50.5%;  Score 633.4;  DB 21;  Length 1505;
Best Local Similarity   71.7%;
Matches  845;  Conservative    0;  Mismatches  331;  Indels   3;  Gaps  1;

Qy     76 AAGCAGGGGGCGGAGGGCCTCGCGGCCCACTACTAATTCTCTCAACATCCACGAGCATCAGGTCCTT 135
Db    208 AAGGCAGGCAGCGGACTCCCGCCAATATTATCGTGTCCAAGATTGAAGGAACTCCAGCTGATT 267

Qy    136 CTTCGCCAAAGACTCATAACCTGTCTCGAGGCTGAGAGAAACGACCTCAAATTCT 195
Db    268 GTGAATGATAAGAGCCAAAACCTCCGGAGCTGACTGCAGAGGACAGAACTAAAGTGCT 327

Qy    196 AGGGTGAGGATGCTGCGCGGAAGAATTACAGCTTCTGCAGGAACCCGGCTTCTTATGTCCGGT 255
Db    328 AAAGTTCGCCCTATTTCGGGAGGAGCTACACGCTGTGCAGGAGGGCTTCCTATGTGGGG 387

Qy    256 GAAGTTGTCAAAGTAATGGGCAGAACAAAAGTCCTTGTCAAGGTTCCACCCGAGAAGGAAAA 315
Db    388 GAAGTAGTCCGGGCCCATGGATAAGAGAAAGTGTTGGTCAAGGTACATCCTGAAGGTAAA 447

Qy    316 TATGTGTTGACATTGACAAAAATATTGACATTACAAAAGATTACTTCCATCCACTAGAGTT 375
Db    448 TTTGTTGTTTAGACGTGGACAAAAAACATTGACATAATGTGACACCCAATTGCCGGGGTG 507

Qy    376 GCACTCCGCAACGACAGTTATGTTCTTCACTTAGTTCTCGCCAGTAAAGTTGATCCATTG 435
Db    508 GCTCTAAGGAAATGAGCTCACACTCCTGACAAGATCCTCCACAAGGTAGACCCATTA 567

Qy    436 GTCAATCTGATGAAAAGTTGAGAAGAGTTCCGGATTCTTCATATGACATGATTGGTGGGTTTA 495
Db    568 GTGTCACTCGATGATGGTCGGAGAAAGTACCAGATTCAACTTATGATGATTGGTGGGACTG 627

Qy    496 GACCAGCAAAATTAAAGRAATAAAAGARAATAAAAGAGTCATTGAGCTACCAATCAAACATCCTGAGGTG 555
Db    628 GACAAACAGATCAAAGGAGAATCAAAGAACTGATCGGAGCTGCCGTGTTAAGCATTCCTGAGGCTC 687

Qy    556 TTTGAAAGTCTTGGAATTGCACACCAAAGAGGGTGTCCTGCTCTATGGGCCACCTGGTTACA 615
Db    688 TTCGAAGCACTCGGGCATTGCTCAGCCCCAAGGGAGTGCTCGTATGGACCTCCAGGGACT 747

Qy    616 GGTAAAACATTGTTGGCGTTAGGGCCGGCTCATCTACTGACTGTACATTCATCCAGGGGTG 675
Db    748 GGGAAGAGACACTGTTGGCCCGGGCTCATCACTACGGACTGTACCTTTATTCGGTCGTC 807

Qy    676 TCTGGGTTCTGAGTTAGTTCAGAAAATACATTGGAAGAAGGTTCTAGAATGGTCAGGGAACTT 735
Db    808 TCTGGCCTCTGAATTGGTTGTACAGAAATTCATCAGGGAAGGCAAGAATGGTGAGGGAGCTG 867

Qy    736 TTTGTTATGGCCAGGGGAACATGCTCCATCCAATTATCTTCATGGATGAAATTGACAGTATT 795
Db    868 TTTGTCATGGACACGGGAACATGCTCCATCACTTTATCTGGATGAAATTGACTCCATC 927

Qy    796 GGATCTGCTCGGATGGAAATCTGGAAGTGGCAACCGGTGGATAGTGAGGTTACAGGGTACTATG 855
Db    928 GGCTCCTTCGGCGGCCTGGAGGGGGGGTTCTGG---AGGGGGACAGTGGAGGTGCACGGATG 984

Qy    856 CTGGAACTTCTCAACCAGTTGGATGGATTGAAGCTTCAAAATAAGATCAAAGGTTTCGATG 915
Db    985 CTGGAGTTGCTCAACACCAAGCTYGACCGGCTTTGAGGGCCACCAAGAACATCAAGGTTATCATG 1044

Qy    916 GCCACCAATCCGGATTGATATCCTGGATCACACCCTCCTTAGACACCAGGACCTCAGTGACCGG 975
Db   1045 GCTTACTAAATAGGAGTTGATTGATGATATCCTGGACTCGGGACTTTGAGGATGGCACTGTCGGGA 1104

Qy    976 AAAATTGAATTTCCAACCCCTTAATAGAAGAGTCTCGCGTGCGATTATTTTGAAAATCCATTCT 1035
Db   1105 AAAATTGAATTCCCACCCCCCAATGAGGAGGCCCGGCTGGACCATTTTGAACATTCATTCT 1164

Qy   1036 AGAAGAATGAATTTAATGCCGTGGCCATTGATTGAAGAAGATTGCCGAGAAGAGATGAATGGA 1095
Db   1165 CGGAAGAATGAAACCTGACCCGGGGGGGATCAACCTGACGAAAAAATTGCTGAGCTCAATGCCAGGA 1224

Qy   1096 GCATCTGGTGGCTGAACGTTAAGGCTGGTTTGCACTGAAGGCTGGAAATGTTTTGCTTTGAGGGAG 1155
Db   1225 GCATCAGGGGCTGAAGTGAAGTGGCCGTGTGCAACAGGAGACATCAGTGTATGCCCTGCCAGAA 1284

Qy   1156 CGGAGGGTACACGTGACTCAGGGAGGGGGATTTTTTGAGATGGCCCTTGGGCGAAGGTGATGAAAAG 1215
Db   1285 CGGCGGAGTCCATGTCACTCAGGAGGGACTTTGAGGATGGCAGTAGCCGAGGTCATGCGGAGAG 1344

Qy   1216 GAGACTGAAAAAAACATTGTCATTGCGGGAAGTTGTGGAG 1254
Db   1345 GACAGTGAGAGAAAAACAATGTCCATCAAGAAAATTATCGGAAG 1383

RESULT 8
ABQ55017
ID   ABQ55017 standard; cDNA; 1800 BP.
XX
AC   ABQ55017;
XX
DT   22-AUG-2002  (first entry)
XX
DE   Human ovarian antigen HSPSG28 cDNA, SEQ ID NO:897.
XX
KW   Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW   ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW   infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW   PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW   inflammatory condition; immune disorder; blood disorder;
KW   cardiovascular disorder; respiratory disorder; neurological disorder;
KW   gastrointestinal disorder; urinary system disorder; neurological disorder;
KW   gene therapy; chromosome mapping; forensic analysis;
KW   antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW   antiinflammatory; gynaecological; reproductive; chromosome 17q23-25;
KW   gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200200677-A1.
XX
PD   03-JAN-2002.
XX
PF   07-JUN-2001; 2001WO-US18569.
XX
PR   07-JUN-2000; 2000US-2094672.
```

XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Birse CE, Rosen CA;
XX
PN   WPI; 2002-147878/19.
DR   P-PSDB; ABP41940.
XX
PT   Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT   useful in the prevention, treatment and diagnosis of cancer (e.g.
PT   ovarian cancer), immune disorders, cardiovascular disorders and
PT   neurological diseases -
XX
PS   Claim 1; SEQ ID No 897; 2922pp; English.
XX
CC   The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC   ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC   encompasses polypeptides 90% identical and polynucleotides 95% identical
CC   to the sequences of the invention. The invention additionally relates to
CC   recombinant vectors and host cells comprising human ovarian antigen
CC   polynucleotides, antibodies against human ovarian antigens, and the use
CC   of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC   treating, prognosing or preventing various ovary and/or breast-related
CC   disorders. Such conditions include ovarian cancer and breast cancer, and
CC   metastatic tumours of ovarian or breast origin, reproductive system
CC   disorders (e.g. infertility, disorders of pregnancy, anovulation,
CC   polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC   disorders, infections (e.g. chlamydia, HIV, toxoplasmosis, and toxic
CC   shock syndrome), inflammatory conditions (e.g. mastitis, oophoritis and
CC   vaginitis), immune disorders (e.g. congenital and acquired
CC   immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC   blood-related disorders (e.g. anaemia), cardiovascular disorders,
CC   respiratory disorders, neurological disorders, gastrointestinal disorders
CC   and urinary system disorders. Ovarian antigen polypeptides and
CC   polynucleotides may also be used in screening for compounds which
CC   modulate ovarian antigen expression or activity. The polynucleotides may
CC   further be used for gene therapy, chromosome mapping, in the
CC   identification of individuals and in forensic analysis, and the
CC   polypeptides may be used as food additives or to prepare antibodies
CC   useful in disease diagnosis, drug targeting and phenotyping. The present
CC   sequence represents cDNA encoding a human ovarian antigen of the
CC   invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1800 BP; 482 A; 404 C; 558 G; 347 T; 9 other;

Query Match   50.5%; Score 633.4; DB 24; Length 1800;
Best Local Similarity   71.7%; Pred. No. 4.8e-186;
Matches 845; Conservative   0; Mismatches 331; Indels   3; Gaps   1;

```
Qy         76 AAGCAGGGCCGAGGGGCCTCCGCCACTACTATTCTCTCAACATCCACGAGCATCAGCTCCTT 135
              ||||| |||||||||||| ||||| ||||||||||||  |||| |||||||| |||||||||
Db        804 GCTCTAAGGAATGACAGCTACACTTCTCCACAGAGATCTTCGCCACAGATCTTGCCAATCA 863

Qy        136 CTTCGCCAAAGACTCATAACCTCAACCGTTCTCGAGGCTCAGGAGAACACACCTCAATTCT 195
              ||||| |||| |||||  |||||||  ||||||||| |||||||||| |||| |||| |
Db        804 GCTCTAAGGAATGACAGCTACACTTCTCCACAGAGATCTTCGCCACAGATCTTGCCAATCA 863

Db        436 GTCAATCTGATGAAAAGTTGAGAAAGTTCCCGATTCTACATATGACATGATCATCAGGGTG 495
Qy        196 AGGGCTGAGGATGCTCGCGAGAGAATTACAGCTTCTCCAGGAGAACCCGCGCTCTTATCTCGGT 255
Db        504 AAGGCAGGCACCGGACTCCGCCAATATTATCTGTCCAAGAGTTGAAGAACTCCAGCTGATT 563
Qy        256 GAAGTTGTCAAAGTTAATGGGCAAGAAGACAAAGTCCTTGTCAAGGTCACCCAGAGGAAAA 315
Db        684 GAAGTAGTCCGGCGGCCATGGATAAGGAGAAGTGTTGGTCAAGGTACATCCTGAAGGTAAA 743
Qy        316 TATGTGTTGTTGACAAAAAATATTGACATTACAAAGATTACTCCATCCACTAGAGTT 375
Db        744 TTTGTTGTTAGACGTGGACAAAAAACATTGACATCAATGAATGTGACACCAAATTGCCGGGTG 803
Qy        376 GCACTCCGCAACGACGAGCAGTTATGTGTTCTTCACTTAGTTCTGCCAAGTAAAAGTTGATCCCATG 435
```

RESULT 9
AAT42343
ID   AAT42343 standard; cDNA; 1287 BP.
XX
AC   AAT42343;
XX
DT   19-FEB-1997 (first entry)
XX
DE   Human 26S proteasome P45 subunit gene.
XX
KW   Human 26S proteasome P45 subunit; primer; PCR; polymerase chain reaction;
KW   amplification; probe; hepatoma cell line HEPG2; antibody; metabolism;
KW   immunodiagnosis; viral infection; ds.

```
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT
FT   CDS             12..1232
FT                   /*tag= a
FT                   /product= 26S proteasome P45 subunit
FT                   1260..1265
FT   polyA_signal    1260..1265
FT                   /*tag= b
XX
PN   (SUME ) SUMITOMO ELECTRIC IND CO.
XX
PA   JP08217796-A.
XX
PR   14-FEB-1995;  95JP-0049103.
PF   14-FEB-1995;  95JP-0049103.
PD   27-AUG-1996.
XX
DR   WPI; 1996-439573/44.
DR   P-PSDB; AAW06100.
XX
PT   Human proteasome P45 subunit and corresp. DNA sequence - useful in
PT   immuno:diagnosis of e.g. Viral infection and metabolic disorders
XX
PS   Claim 2; Page 3; 21pp; Japanese.
XX
CC   This is the nucleotide sequence encoding the P45 subunit of the human
CC   26S proteasome complex. The gene was isolated using primers AA142344-5
CC   which amplified a 600 bp fragment of the coding sequence. The primers
CC   were generated based on peptide sequence of the purified protein. The
CC   amplified fragment was used as a probe to isolate the full length
CC   sequence from a cDNA library derived from mRNA purified from human
CC   hepatoma cell line HEP62. The protein can be used to raise antibodies
CC   for immunodiagnosis of diseases such as viral infections and abnormal
CC   metabolism.
XX
SQ   Sequence 1287 BP; 364 A; 276 C; 380 G; 267 T; 0 other;

Query Match     49.9%;  Score 625.4;  DB 17;  Length 1287;
Best Local Similarity   71.2%;
Matches  840;  Conservative   0;  Mismatches  336;  Indels   3;  Gaps   1;
                                       Pred. No. 1.2e-183;
```

```
Qy        76 AAGCAGGGGCGGAGGGCGCCTCCGCCACTACTATTTCTCAACATCCACGAGCATCAGTCCCTT 135
             |||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db        54 AAGGCAGGGCAGCGGCGGGACTCCGCCAATATATTCTGTCCCAAGATTGAAGAACTCCAGCTGATT 113

Qy       136 CTTCGGCCAAAAGACTCATTACCTCACCGTCTCTGGAGGCTCAGAGAAACGACCTCAATTCT 195
             ||||  |||||||  ||||||||||  ||||||||||||||||||||||||||||||  ||
Db       114 GTGAATGATAAGAGCCAAAAACCTCCGGAGGGCTGCAGCGCACAGAGGAACGAACTAAATGCT 173

Qy       196 AGGGTGAGGATGCTGCCGGGAAGAAATTACAGCTTCTTCGCGAGGAACCCGGCTCTTATGTCGGT 255
             ||  |||||||||||||||||||||||||||  |||||||||||||  |||||||||||||
Db       174 AAAGTTCCGCCTATTGCGGAGGGAGCCTACAGCTGCCGGGGCTCCTTATGTGGGG 233

Qy       256 GRAGTTGTCAAAGTAATGGGCAAGAACAAAAGTCCTTGTCAAGGTCCACCCAGAAGGAAAA 315
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       234 GRAGTAGTCCGGCGCCATGGATAAGAGAAAAGTGTTGGTCAAGGTACATCCTGAAGGTAAA 293

Qy       316 TATGTTGTTGACATTGACAAAAAATATTGACATTACAAAGATTACTCCATCCACTAGAGTT 375
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       294 TTTGTTGTTAGACCTGGACAAAAACATTCACATCAAATGATGTGACACCCAATTGCCGGGTG 353

Qy       376 GCACTCCGCAACCACCAGTTATATGTTCTTCACTTAGTTCTGCCAAGTAAAGTTGATCATGG 435
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       354 GCTCTAAGGAATGACACGCTACACTCTCGCACCAAGATCCTGCCCAACAAGGTAGACCCATTA 413

Qy       436 GTGTCACTGAAGAATGACAGCTACACTCTCGCACCAAGATCCTGCCCAACAAGGTAGACCCATTA 495
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       414 GIGTCACTGAAGAAAGTACCAGATTCAACTTATGAGATGATTGGTGACTGG 473

Qy       496 GACCAGCAAATTAAGAAATTAAAGAGAGTCATTGAGCTACCAATCAAACATCCTGAGCTG 555
             ||||||||||||||||||||||||||||||||||||||||||
Db       474 ACAAAACAGAATCAAGGAGATCAAAGAGAAGAGAAGTGATCAGCTGCCTGTTAAGCCACCTG 533

Qy       556 TTTGAAAGTCTTGAAATTGCACAACCAAAGGGTGTCCTGCTCTTATGGGCCACCTGGTACA 615
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       534 TTCGAAGCACTGGGCCATTGCTCAGCCCACAAGGGAGTGCTGCTATTGCACCTCCAGGCACT 593

Qy       616 GGTAAAAACATTGTTGGGATCTAGGGCAGTAGGGTTCATCTATACATTACATTCATCAGGGTG 675
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       594 GGGAAGACACTGTTGGCCCGGGCGTGTGGCTTCATCTATACTGACTGACAGTATT 653

Qy       736 TTTGTTATGCCAGGACAACATGCTCCATCAATTATCTTCACTGATGAAATTGACAGTATT 795
             |||||||||||||||||||||||||||||||||||||||||||||
Db       654 TCTGGCCTTCGAAATTGGGTACAGAAAATTCATAGGGAAGGGCCAAGAGCTG 713

Qy       796 GGATCTGCTTCGGGATGGAAATCTGGAAGGTGGCCAACGGGTGGATTAGGTGAAGGTACAGTGG 855
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       714 TTTGTCATGGCACGGGAACATGCTCCATCAATCATCTTCATGGGACGAAATCGACTCCATC 773

Qy       856 CTGGAACTTCTCAACCACAGTTGGATGGATTGGAAGGTTCAAATAAGATCAAAGGTTTGATG 915
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       774 GGCTCCTTCGGCGGCTGGAGGGGTTCCTGG---AGGGGACAGTGAAGTGCAGCCGCACGAATG 830

Qy       916 GCCACCAATCGGATTGATTGATATCCTGGATCAAAGCCCCTCCTTAGACCAGGACGACACGG 975
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       831 CTGGGAGTTGGCTCAACCCAGCTCGACGGCCTTTGAGCGCCACCCAAGGAACATCAAAGGTTATCATG 890

Qy       976 AAAAATTGAAATTCCAACCCCTAATGAAGACTCTCGGCTGGATATTTTGAAAATCCATTCT 1035
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       891 GCTACTAATAGGATTGAATATCCCTGGACTCCGGCACTGCTTCGCGCCACCATTGACACAGA 950

Qy      1036 AGAAGAAATGAAATTTAAATGGTGGCATTGATTTGAAGAGATTGCCGAGAAGAATGGA 1095
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       951 AAAATTGAAATCCCACCCCCAATGAGGAGGCCCGGCTGGACATTTGAAGAATCCATTCT 1010

Qy      1156 CGGAGGGGTACACGTGACTCAGGAGGATTTTGAGGATTGGCCGTGGCGGAAGGTGATGAAAAG 1215
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1131 CGGCGAGTCCATGTCACTCAGGAGGACTTTGAGGATTGGCAGTGGCATCAGAGAAG 1190

Qy      1216 GAGACTGAAAAAAACATGTCATTGCGGAGTTGTGGAAG 1254
             |||||||||||||||||||||||||||||||||||
Db      1191 GACAGTGAGAAAAACATGTCCATCAAGAAATTATCGGAAG 1229
```

```
RESULT 10
ABL02721
ID   ABL02721 standard; cDNA; 1396 BP.
XX
AC   ABL02721;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 2645.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
OS   Drosophila melanogaster.
XX
PN   WO20017042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
```

PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB58618.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 2645; 21pp + sequence listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1396 BP; 375 A; 335 C; 387 G; 299 T; 0 other;

Query Match   48.5%; Score 608.4; DB 23; Length 1396;
Best Local Similarity   70.2%;
Matches 832; Conservative   0; Mismatches 351; Indels 3; Gaps 1;
                          Pred. No. 2.5e-178;

Qy   69 GCCCACCAAGCNCAGGGGCGGCGGGGCCGGGTCCCGCCCACTACTATTCTCANCATCCACGAGCATCA   128
Db   151 GGCCTACCATAAAGGCGGAGGGATTCCGTCCTTACTACATCTCAAAAGATCGAGGAACTGCA   210

Qy   129 GCTCCTTCTTCGGCCAAAAGACTCATAACCTCAACCGTCTCGAGGCTTCGAGAAAACGACCT   188
Db   211 GCTCGGTGGTGGGCCGGAGAGCACCAGAAATTTGAGACGACTGCAGGCCCCCAGCGCAATGAGCT   270

Qy   189 CAATTCTAGGGTGAGGATGCTCGCGAAGAATTACAGCTTCTGCGACGAACCCGGGCTCTTIA   248
Db   271 CAATGCTAAAGTTCCGGATGCTCGCGGAGGAGCTGCACCTGTCTCCAGACAACAGGGTAGCTA   330

Qy   249 TGTCGGTGAAGTTGTCAAAGTAATGGGGCAGAACAAAGTCCGGATTCTGTCAAGGTCCACCCGA   308
Db   331 TGTGCGGCCAGGTTGTGAAGCCCATGCACGAAGAAAGTGCTGGTCCAAGGTTGCCATCCGA   390

Qy   309 AGGAAAATATGTTGTTGACATTGACAAAAATATTGACANGATTACTTACCATCCAC   368
Db   391 GGGCAAGTTCGTGCTCGGATCTGGACAAAAACATTAGATATCAACGACGTAACCCCAAATG   450

Qy   369 TAGAGTTGCACTCCGCAACGACACAGTTATTGTTCTTCACTTAGTTCTTCCCAACGTAAAGTTGA   428
Db   451 CCGCGGTGCCCTGCGCAATGAGACTATTACGCTGCACAAAGATTCTGCCCAATAAGGTGGA   510

Qy   429 TCCATTGGTCAATCTGATTGAAGAAGTTGAGAGAAGTTCCGGATTCTACATATGCATGATTGG   488
Db   511 TCCGCGCTGGTCTCCGCTTATGATGGTTGAAAAAGTTCCTGACTCCACATACGTAAATCGTTGG   570

Qy   489 TGGTTTAGACCAGCAAAATTAAAGAAAATAAAAGAGGTTCATCGAGGCTACCAAATCAAACATCC   548
Db   571 CGGCCTGGACAAAACAAGATTAAAGACAGAGGTGATTGAGTTGCCCGTAAAGCATCC   630

Qy   549 TGAGCTGTTTGAAAGTCTTGGAATTGCACAACCACCAAAAGGGGTGTCCTGCTCTATGGGCCACC   608
Db   631 AGAGCTGTTTGAATGCCTTGGGTTATTGCCCAGCGAGGGAGTGGCTCCCAGAAGGGCATCTC   690

Qy   609 TGGTACAGGTGTAAAAACATTGTTGTTGGGCTAGGGGCGGGGGAGTGCAC   668
Db   691 AGGGTACAGGAGAAAGAGACTTTGCTGGCGGAAGACTGACTGTACAGACTTCAT   750

RESULT 11
ABL04117
ID   ABL04117 standard; cDNA; 1318 BP.
XX
AC   ABL04117;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 6833.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US092931.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB60014.
XX

Qy   669 CAGGGGTGTCTGGGTTCTGAGTTAGTTCAGAAATACATTGGAGAAGGTTCTAGAATGGTCAG   728
Db   751 TCGCGGTCTCTGGCTCAGAGACTGGTTCCAGAAAATTCAATCGGCGAGAGGTTCGCGCATGGTTCG 810

Qy   729 GGAACTTTTTGTTATGGCCCAGGGAGAACATGCTCCATCAATTATCTTCATGGATGAAATTCA   788
Db   811 AGAACTCTTTGTTAATGGCCCGGGAAACATGCACCATCAATTATTTTTCATGGACGAAATCGA   870

Qy   789 CAGTAATTGGATCTGCTCCGGATGGAATCTTGGAAGTGGCAACGGTGATGAGTAGTACAGCG   848
Db   871 TTCCATTGGCTCCGTCGGCGTTGATTCAGTC---AGGTTCCGGCGCGATTCCGAGGGTGCAGCG   927

Qy   849 TACTANGCTGGGACTTCTCAACCAGTTGGATGGATTGAAGCTTCAAATAAGATCAAGGT   908
Db   928 TACTATGCTGGGACTCACTTACACCAGCTGGACGCGGCTTTGAGGCCACCAAGAAACATCAAGGT   987

Qy   909 TTTCGATGGCCAACAATCGGATTGATATCCTGGATTCAAGCCCTCCTTAGACCAGGACGGAT   968
Db   988 GATCAGGCCCACCAATCCCATTGAACATCCTGGAATCCCGGCTCTGCCTGCCTCCCGGCGTAT   1047

Qy   969 AGAACCGGAAAAATTGAATTCCAACCCTAAATGGAAGTTCCGGCTCTGGATTATTTTGAAAAT   1028
Db 1048 TGATCGCAAGATCGAGTTCCCGCCACCAAACGAGGAGCGCGCGTCTGGACATCTTGAAGAT   1107

Qy 1029 CCAATTCTAGAGAATGAAATTTAATGCCGTGGCATTGAATTTGAAGAAGATGTGCCGAGAAGAT   1088
Db 1108 CCACTCCCGTAGCATGAAACCTCACGCCTGGCATTAATCTGCCAAGAATCGCCGAAGCTAAT   1167

Qy 1089 GAATGGAGCATTCGGTGCTGAACTTAAGGCTGTTGCACTGAAGTCTGAAATGTTTGCTTT   1148
Db 1168 GCCAGGCCCATCAGGTGCGGGAAGTCAAGGGCGTCTGCACGAGGCCCGCCGCATGTATGCCCT   1227

Qy 1149 GAGGGAGGGGGAGGGTACACGTGACTCAGGAGGGGAGGATTTTGAGATGGGCCGTGGGCAAGGTGAT   1208
Db 1228 GCGGGAAGCGACGGCGTCCCATGTCACCCAGGAGGAGGACTTCGAGGATGGGCTGTGGCCAAGGTCAT   1287

Qy 1209 GAAAAAGGAGACTGAAAAAACATGTCATTGGGGAAGTTGTTGGGAAG   1254
Db 1288 GCAGAAGAGACTCGGAGGAAGAACATGTCTATCAAGAAGCTGTGGGAAG   1333

```
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 6833; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent,
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signaling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL6176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57337-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1318 BP; 371 A; 286 C; 359 G; 302 T; 0 other;

Query Match          46.0%;  Score 576.8;  DB 23;  Length 1318;
Best Local Similarity 69.0%;  Pred. No. 1.7e-168;
Matches   807;  Conservative    0;  Mismatches  357;  Indels    6;  Gaps    1;

Qy    84 CAGGGGGCCTCCCGCCACTACTATTCTTCAACATCCACGAGCGATCAGCCTCCTTCTTCGCCA 143
         |||  |||| |   |||   |     |||     |      || |  || |||   || |||||
Db   151 CGAAGGCTTTCATTCATATATTATACGCAGAGAAAATATCAGAGCTTCAGTTCACTGTGAACGA 210

Qy   144 AAAGACTTCATAACCTCAACCGTCTCGAGGCTCAGAGAGAAACGACCTCAATTCTTAGGGTGAG 203
         |||||       | |||||||  | |||||| || || || ||||||| |||| |||| |||||
Db   211 ACCCCAAAAAGAACTTGCTCCGTTTGCAGGGCTCAGAGAAAATGAAACTCAAACCTCAAGGTCG 270

Qy   204 GATGCCTCCGCGAAGAATTTACAGCTTTCTTGCAGGAAACCCGGCTTTATTGTGTCGGTGAAGTTGT 263
         |||| | |||  |    || |   | |||||  |||| || |||||| |  || |   |||||||||
Db   271 CTTGGCTGCGCGAGGAGCTGCAGTTGCTCCAGGAGGACTATATTGCCGATGTGT || |||| 330

Qy   264 CAAAGTAATGGGCAAGAACAAAGTCCTTGTCAAGGTTCCACCCAGAGAGGAAAATATGTTGT 323
         |||||   ||||  ||    | |  ||| ||| || |||  ||| | || || || ||||||
Db   331 GAAGCCCATGGACAAAGAATAAGGTCCTTGTGAAGGTTCATCCCGGAGGGCAAGTAGTGAGT 390

Qy   324 TGACATTGACAAAAATATTGACATTACAAAGATTACTCCATCCACTAGAGTTGCACTCCG 383
         ||||| |||||  | |||||| |||| || |||||| |||| ||||  |||||||||||
Db   391 GGATGTGGACAAGACCATCAATATATTAAGGATGTGACACCCAGCAGGGGTGGCCCTGCG 450

Qy   384 CAACGACAGTTATGTTCTTCACTTAGTTCTGCCAAGTAAAGTGATCAGTTGGTCAATCT 443
         |||||  |||||| |  || |||| |||| | |||||  |||| ||||| |||| || |
Db   451 AAACGAGAGCTACACCGCTTCACAAAATTCTGCCCAATAAGGTGGGACCCATTAGTATCCCT 510

Qy   444 GATGAAAGTTGAGAAAGTTCCCGATTCTTACATATGACATGATTGGTGGGTTTAGACCAGCA 503
         |||| |||| | |||| ||   ||| |  |||| || || | |||  |||| |||| |||||
Db   511 TATGTTGGGTGGAAAGGTGCCCGACTCCACCTTACGGTTTGGGCTGGCTTGGACAAGCA 570

Qy   504 AATTAAGAAAATPAAAAGAGGTCATTGAGCTACCAAATCAAACATCCTGAGCTGTTTGAAAG 563
         |||||   |||     |||| |||  || |||  ||| |||| |||| |||| ||||| |||
Db   571 GATCCAGAGATCAAGGAGGTGATTGAGCTGCCTGTAAAGCATCCCGGAATTATTTTGACGC 630

Qy   564 TCTTGGGAATTGGCACAACCAAAAGGGGTGTCCGTGCTCTGATGTGGGGSCCACCTGGTACAGGTNAAAAC 623
         ||||||  |||| || | ||||  |||| || |||| | | |||  || |||| ||||  || |  ||||
Db   631 CTTGGGCATTACCCAGCCAAAGGGAGTGCTCCTCGACCGCCTGGAGCTGGGAGCCTGTACAGGTGAAAAC 690

Qy   624 ATTGTTGGCTAGGCGAGTGGCCTCATCATTACTGACTGTACATTCATCAGGGTGTCTGGTTC 683
         |||| |   ||| |  |||  | || |  |||  | | || |||||   |||| ||  ||||
Db   691 CCTGTTAGCCCAGCTGCTCACACGGAGGTCACTTTTTACGGTTTCAGGGTC 750

Qy   684 TGAGTGTAGTTCAGAAATACAATTGGAGAAGGTTCTAGAAATGGTTCAGGGAACTCTTTTGTTAT 743
         ||||| | ||||| |   ||  ||  || || | ||||| ||  || |  |  ||||||||| |||
Db   751 GGAACTGGTTCAGAGATTCATCGGGCAGGGTTCGGAGAATGCTGAGGAGNTGTAAGGGAT 810

Qy   744 GGCCAGGGAACATGCTCCATCAAATTACTTCATGGATGAAATTGACAGTTAATTGACAHTGC 803
         ||||| || | || ||  ||||  |||| ||||  ||  || |  || |||| |  | ||||
Db   811 GGCTAGGGAACACGCTCCATCCCATAATATTATTCATGGACGAGGATTCGGAGTTCGATTTGGTTCCGC 870

Qy   804 TCGGATGGAAGTCGGAAGTGGGCAACGGATCGATGAGTGAGGTACAGCCGTTACTAGTGCTGGAACT 863
```

```
Db   871 TCGACTGGGAAACCCGA------ACGGGCGGATTCGGAGGTCCAGCAGACCATGCTGGAGCT 924
         ||||  ||||||| |||       ||||  |||||||||||||||||||||||||||||||
Qy   924 TCGGAATTGATATCCTGGAATTTGAGAGCTTCAAATAAGATCAAGGTTTTGATGGCCACCAA 923
         |||   ||||| |  |  | ||| |  |||||  ||  | | || |||  | ||||| |||||
Db   925 TCTGAACCAACTGGACACGGCTGCTTCGAGCGCTACTAAAAACAATTAAGGTGATCATGGCCACCAA 984

RESULT 12
ABL02720/c
ID   ABL02720 standard; cDNA; 3604 BP.
XX
AC   ABL02720;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 2642.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB58617.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 2642; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent,
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
```

```
Qy  1104 TGCTGAACTTAAGGCTGTTGCACTGAAGCTGAATGTTTGCTTTGAGGGAGGCGGAGGT 1163
         |||||| | |||||  || |||| | |||  ||||| |||| | |||||| || ||||
Db  1165 TGCTGAAGTCAAGGGGTGTTTGCACCGAGGCGGGAATGTATGCTCTGAGGGAAAGACGGGT 1224

Qy  1164 ACACGTGACTTCAGGAGGATTTTGAGATGGCCGTGGGCGAAGGTGATGAAGAGGAGACTGA 1223
         | | |||||  ||||||| || ||||||||||| |   ||  ||||   || || || |||
Db  1225 CCATGTCACCCAGGAGGACTTTGAAATGGCCGGTCTCACAGGTAAATGAATGAGGACTCCGA 1284

Qy  1224 AAAAAACATGTCATTGCGGAAGTTGTGGAA 1253
         ||   ||| ||| | | |||| ||| ||||
Db  1285 GAAGAACAATGTCCATCAGAAAGTTCTGGAA 1314
```

Left column

```
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC   sequences (ABL16176-ABL30511), and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3604 BP; 969 A; 830 C; 803 G; 1002 T; 0 other;

Query Match      45.0%;  Score 563.8;  DB 23;  Length 3604;
Best Local Similarity   71.3%;
Matches  758;  Conservative    0;  Mismatches  302;  Indels    3;  Gaps   1;

Qy   192  TTCTAGGGTGGAGGATGCTGCGCGAAGARATTACAGCTTCTGCAGGAACCCGGCTCTTATGT  251
          |||||||| ||||||||||| ||||   |||||||||| || ||||||||| |||  | ||||
Db   2123 TTTTGCAGTTCGGATGCTGCCGGGAGGAGCTGCAGCTGCCTCAGGAACGACAGGGTGGTATGT  2064

Qy   252  CGGTGAGTTGTCAAAGTAATGGGCAGGAACAAAAGTCCTTGTTGCAGGTGTCACCCAGGAAGG  311
          |||||| ||| || |||||||  |||||| || || |||| || |||||||||||||| || |
Db   2063 GGGCGGAGGTTGTGAAGCCCATGGACAACGAAAGTGCCTGGTTCAAGGTCATCCCAGGGG  2004

Qy   312  AAAATATGTTGTTGACATTGACAAAAAATTGACATTACAAAGATTACTCCATCCACTAG  371
          |||| |||  || |||||||| ||||  |||| |||| || || |  | || |  | || 
Db   2003 CAAGTTCTTCGTCGGATCTGGACAAAAACATAGATATCAACGACGTAACCCCAAATTGCCG  1944

Qy   372  AGTTGCACTCCGCAACGACCAGTTAGTTGTTCTTCACTTAGTTCTGCCAAGTAAAGTTGATCC  431
          ||| |  |||| |||  ||| |||   | |  |||| || |  ||||| || || ||||||| 
Db   1943 CGTGGCCCTCGCCAATGAGAGCTATACCGTCGCACAAGATTCTGCCCAATAAGGTGGATCC  1884

Qy   432  ATTGGTCAATCTGATGAAAGTTGAGAAGTTCCCGATTCTACATATGCACATGATTGGTGG  491
          ||| |||||| |||| ||||||||| |||||||| ||| |||| |||| |||| |||||| |
Db   1883 GCTGGTCTCGCTTATGATGGTTGAAAAGTTCCTGACTCCACGATACGAAAATGGTTGGCGG  1824

Qy   492  TTTAGACCAGCAAATTAAAGAATAAAGAGGTCATTGAGCTACCAATCAAACATCCTGA  551
          |||| ||||  | |||||  |||| | | ||||| || ||||| ||||  |||| || 
Db   1823 CCTGGACAAACAGATTAAAGAGATCAACGAGGTTGATTGAGTTGCCCGTAAAGCATCCAGA  1764

Qy   552  GCTGTTTGAAAGTTGCACACACCAAAGGGTGTCCTGCTCTATGGGCACCTGG  611
          |||| | |||| || ||||| |  || | |||| || || ||  | ||||| 
Db   1763 GCTGTTTGGATGCCCTGGGGTTATTGCCCACGGCGAGGGGAGGTGCTCCCTTT  1704

Qy   612  TACAGGTAAAACATTGTTGGCTAGGGCGTCATCATACTGACNGTACAATTCAATCAG  671
          ||||| || ||| |||| |||  |||| ||| || || |  |  | ||||| ||||| 
Db   1703 TACAGGAAAAGACTTTGCTGCGGCCAGGGCCCGTTGCCCATCACTACACCGAGTGCACCTTCAATTCG  1644

Qy   672  GGTGTCTGGTTCTGAGTTAGTTCAGAAATACATTGGAGGAAGGTTCTTAGAATGGTCAGGGA  731
          |||||| || || || | ||| |||| | |||  | || |||| || |  || ||| |||| 
Db   1643 CGTCTCTGGCTCCAGGCTGGTCCAGAAANTCACNTCGGCCGCAGGGGTTCCGGGCATGGTTCGAGA  1584

Qy   732  ACTTTTTGTTATGGCCAGGGAACATGCTCCATCCATATATCTTCATGGGATGAGAAAATTGACAG  791
          ||||||||  ||||||| |||||    |||||| | || |||| ||| | | ||||| ||||| |
Db   1583 ACTCTTTGTGAATGGCCCCGGGGAACACGCACCATCAAATTATTTTCATGGACGAGAAAATCGAGTTC  1524

Qy   792  TATTGGATCTGCTCGGATCGGAAATCTGGAAGTGGCAACGGGTGATAGTTGAGGTTACAGCGTAC  851
          ||||||| || ||  |||  |   | |||| ||||| || ||  ||| | || ||||||||||||
Db   1523 CATTGGCTCGGTCGGCGGTTATTGAGTC---AGGTTCCGGCGGCGANTTCCGAGGGTCCAGCGGTAC  1467

Qy   852  TATGCTGGAACTCTCACACCAGTGGATGGGATTTGAAGCTTCAAATAAGACATCAAGGTTTT  911
          || |||| ||| |  |||||| ||| || ||||  ||||| |||| ||| |||||| ||||
Db   1466 TATGCTGGAGCTACTAACCAGCTGGATTTTGAGGCCACCAACGAACATCAAGGTGAT  1407

Qy   912  GATGGCCACCAAATCGGATTGAATTGAATATCCTGGATCAAGCCCTCCTTAGACCAGGACGGATAGA  971
          ||||||| ||||| ||||||||| ||| |||| |||||  | || || |  ||  |||| || | |
Db   1406 CATGGCCACCAAATCGCACATCCGCGTCTGTGCCGTGCGGTCGTCGGCAAGAACATCAAGGTGAT  1347

Qy   972  CCGGAAATTGAATTCCAACCCCTAATGAACAGNGTCTCGGCTCGGCCATAATTTTTGAAAATCCA  1031
          |||| |||  | ||||| ||| | || |  || ||| ||||| || || |   |||| |||||||
Db   1346 TCGCAAGATCGAGTTCCCGCACCAAAACAGGGCGCGGTCGGCTGACATCTTGAAGATCCA  1287

Qy   1032 TTCTAGAAGAATGAAATTTAAATGGCGTGGGCATTGAATTGAAAGAGGAATTGCCGAGAAGRATGAA  1091
          |||||||| || |   |||| ||| |  ||| |||| ||||| ||||| || || || |  | |||
Db   1286 CTCCCGTAAGAGATGAAACCTCACCGCCGGGTGCGTGGCAAGATGCGGCAAGAGCTAAGTGCC  1227
```

Right column

```
Qy   1092 TGGAGCATCTGGGTGCTGAACTTAAGCTGTGTTTTGCACTGAAGCTGGAAATGTTTTGCTTTGAG  1151
          ||| ||| || || || || |||  || || || ||||| |||  | |||| |  ||| ||||||
Db   1226 AGGCCCATCAAGGTGCGGAAGTCAAGGCGGCGTCTGCACGGAGGGCGCCGCATGTATCCGCTCCG  1167

Qy   1152 GGAGCGGAGGGTTACACGTGACTCCAGGAGGATTTTGACGATGCGGGCGGGGAAGGTGATGAA  1211
          ||| ||| || | |  ||||| | ||||| ||||| || |||||  ||| || || |  | ||
Db   1166 CGAACGACACCGTCCATGTCACCCAGGAGGACTTCGGAGAATGGCTGTGGCCAAGGTGATGCA  1107

Qy   1212 AAAGGAGACTGAAAAAAACACATGTCATTGCGGAAGTGTGTGGGAAG  1254
          |||| | | || || || || | | | || ||| || | | |||||
Db   1106 GAAGGAACTCGGAGAAGGAACAATGTCTATCAAGAAGCTGTGGAAG  1064
```

```
RESULT 13

ABL04116/c

ID   ABL04116 standard; cDNA, 3373 BP.
XX
AC   ABL04116;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 6830.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-1916379P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB60013.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 6830; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC   sequences (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3373 BP; 871 A; 809 C; 711 G; 982 T; 0 other;

Query Match      43.2%;  Score 541.4;  DB 23;  Length 3373;
Best Local Similarity   70.1%;
Matches  744;  Conservative    0;  Mismatches  311;  Indels    6;  Gaps   1;

Qy   193  TCTAGGGTGGAGGATGCTGCCCGAAGAATTACAGCTTCTGCAGGAACCCGGCTCTTATGTC  252
          |  | | |||| || ||||| | || ||  ||||  |||| ||||   | |||| | ||  
Db   2059 TCCACAGTTCGGCTTGCTGCTGCGGGAGGAGCTGCAGTTGCTCCCAGAGGAGAGCTATATT  2000
```

```
RESULT 14
AB232038
ID    AB232038 standard; DNA; 1206 BP.
XX
```

(Sequence alignment — Qy = Query, Db = Database)

```
Qy    253  GGTGAGAGTTGTTGTCAAAGTAATGGGCAGAGAACAAAAGTCCTTGTCGTCAAGGTCCACCCAGAAGGA 312
Db   1999  GCCGGAGGTTGTGHAGCCCATGGCAAAGTCCATTGACATTGAGCCCATGGATTCATCCCGAGGGC 1940
Qy    313  AAATATGTTGTTGACATTGACAAAAATATTGACATTACAAAGATTACTCACTACTAGA 372
Db   1939  AAGTATGTAGTGTGGATGTGGACABAGACCATCAATATTAATATTAAGGARGTGACACCCAGCAGCAGG 1880
Qy    373  GTTGCACTTCCGCAACGACAGTTATGTTTCTTCACTTAGTTCTTCTGCCAGTAAAAGTTGATCCA 432
Db   1879  GTGGCCCTGCGAAACGAGAGAGCTACACGCCTTCACCAAAATTCTGCCCAATAAGGTGGACCCA 1820
Qy    433  TTGGTCAATCTGATGAAAGTTGAGAAAGTTCCCGATTCTACATATGAATGATGATTGGTGGT 492
Db   1819  TTAGTATCCCTTAATGTTGGTGGAAAACGTGCCCGACTCCACCTACGAAAATGGTGGGTGGGC 1760
Qy    493  TTAGACCAGCAAAATTAAAAGAAATTGGAACAGCTTACCCAATCAAACATCCTGAG 552
Db   1759  TTGGACAAAGCAGATCAGCAGGAGTCAAGGAGGTGATTGAGCTAGGCATCCGGAA 1700
Qy    553  CTGTTTGAAAGTCTTTGGAATTGCACACAACCAAAAGGGTGTTCCTCGCTTCTCTATGTGGGCCACCTGGT 612
Db   1699  TTAATTGACGCCTTGGGCATTACCCGCCAAAAGGGAGTGCTCCCTCTACCGACCGGCCTGGA 1640
Qy    613  ACAGGTAAAACATTGTTGGCTAGGGGCAGTGGCTCATCATACTGTACATTCACATCAGG 672
Db   1639  ACTGGGGAAAAACCCTGTTTAGTCCAGCTGTTAGCCCATCACACGAGGTGGCACTTTTATACGG 1580
Qy    673  GTGTCTCGTTCTCAGTTAGTTGTCAGAAATACATTGGACAAGGTTCTAGAAATGAAGTCAGGGAA 732
Db   1579  GTTTCAGGGTCGGGAACTGGTTCAGAAATTCATCGGCCGAGGGTTCGAGAATTGTAAGGGAG 1520
Qy    733  CTTTTTTGTTGGATGGGGCTAGGGAACACGGCTCCATCCATAATAATTCATCATGGATGAAAATTGACAGT 792
Db   1519  CTTTTTTGTTTGATGGGCTAGGGAACAACGGCTCCATCCATAATTCTTCATCATGGACGAGGATCGACTCG 1460
Qy    793  ATTGGATCTGCTCGGATGGAAGTGGCCAACCGGTGATAGTGAGGTACAGCGTACT 852
Db   1459  ATTGGTTCCGCTCGACTGGAAAACCGGA------ACGGGGATTCGGAGGTCCAGCGAACC 1406
Qy    853  ATGCTTGGGAACTTCTCAACCAGTTGGATGGGATTGGAAGCTTCAAAATAAGATCAAGGTTTTTG 912
Db   1405  ATCCTGTGAGCTTCTGACCAACTCTGGAGGGCTTCGAGGGTCTACTAAAAACATTAAGGTGATC 1346
Qy    913  ATGGCCACCAAATGGATTGATATCCTGGAATCAAGCCCTCCTTAGACCCAGGACGGGATAGAC 972
Db   1345  ATGGCCACCAAATCGCATCGGATCGTTCTGGGATTCAGGCATTACTGCGGTCCGGATTGAC 1286
Qy    973  CGGGAAAATTGAATTTCCAACCCCTAATGAAGAGTCTCGGCCTGGATAATTTTGAAAATCCAT 1032
Db   1285  CCGAAGATCGAGAGTTTCCGCCACACCAAACGAAAGGGCCAGATTGGGATAATCCTGAAAGATAACAT 1226
Qy   1033  TCTAGAAGAAATGAAAATTTAATTTTAATGCGTGGGCCATTGAATTGTTGAAAGAAGAGATGAAT 1092
Db   1225  TCTCGCAAGATGAAGAACCTCACCAGGGCCATCAATCTTCGCCAAGATTGCTGAAGAAATGCCG 1166
Qy   1093  GGAGCACTCTGGGTGCTGAACTTAAGGCCGTGTTTGCACTGAAGCTGGBAATGTTTGCTTTGAGG 1152
Db   1165  GCGGGCCTTCTCGGGTGCTGAAGTCAAAGGGTGTTTGGCACCGGAAGTATGTCTCTGAGG 1106
Qy   1153  GAGCGGGAGGGTACACGTGACTCAGGAGGATTTTGAGGATGAGGCCGTGGCGAAGGTGATGAAA 1212
Db   1105  GAAAGACGGGGTCCATGTCACCGGAGGACTTTGAAAATGGCCGTCTCCAAGGTGAATGGATG 1046
Qy   1211  AAGGAGACTGAAAAAAACATGTCAATTGCGGAGAGTTGTTGGAA 1253
Db   1045  AAGGACTCCGGAGAAGAACATGTCATACAGAAAGTTCTGGA 1005
```

(Right column — patent bibliographic fields)

```
AC    AB232038;
XX
DT    30-JAN-2003   (first entry)
XX
DE    Candida albicans essential gene SEQ ID NO 6325.
XX
KW    Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW    signal transduction; DNA replication; cell division; growth;
KW    proliferation; Candida albicans; fungicide; antifungal; gene; ss.
XX
OS    Candida albicans.
XX
PN    WO200253728-A2.
XX
PD    11-JUL-2002.
XX
DR    26-DEC-2001; 2001WO-US49486.
XX
PF    29-DEC-2000; 2000US-2591128P.
PR    20-FEB-2001; 2001US-0792024.
PR    22-AUG-2001; 2001US-3140950P.
XX
PA    (ELIT-) ELITRA PHARM INC.
XX
PI    Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
XX
DR    WPI; 2002-566694/60.
DR    P-PSDB; ABP73488.
XX
PT    Constructing strains for identifying gene products as effective targets
PT    for therapeutic intervention, by inactivating in the strain one allele
PT    of a gene and placing other allele of the gene under conditional
PT    expression -
XX
PS    Claim 37; SEQ ID NO 6325; 167pp + Sequence Listing; English.
XX
CC    The invention relates to constructing (MI) a strain of diploid fungal
CC    cells in which both alleles of a gene are modified, comprising modifying
CC    one allele by insertion or replacement by a cassette having an
CC    expressible selectable marker and modifying other allele by
CC    recombination, of a promoter replacement fragment with a heterologous
CC    promoter, so that expression of the second allele is regulated by the
CC    promoter. (MI) is useful for constructing a strain of diploid fungal
CC    cells in which both alleles of a gene are modified. The diploid fungal
CC    cells having both alleles modified are useful for identifying a gene that
CC    is essential to the survival or growth of a fungus, a gene that
CC    contributes to the virulence and/or pathogenicity of a fungus, a gene
CC    that contributes to the resistance of a diploid fungus to an antifungal
CC    agent, an antifungal agent that inhibits the growth of a diploid fungus
CC    and for identifying a therapeutic agent for treatment of a mammalian
CC    disease. (MI) is useful for identifying a compound which modulates the
CC    activity of a gene product, preferably enzymatic activity, carbon
CC    compound catabolism, biosynthetic, transporter, transcriptional,
CC    translational, signal transduction, DNA replication and cell division
CC    activity. The method is useful for identifying a compound having the
CC    ability to inhibit growth or proliferation of C. albicans cells and for
CC    treating infection by C. albicans. The present sequence is that of an
CC    essential Candida albicans gene used in the method of the invention.
CC    Note: The sequence data for this patent is not represented in the printed
CC    specification but is based on sequence information supplied to Derwent by
CC    the European Patent Office.
XX
SQ    Sequence 1206 BP; 425 A; 170 C; 259 G; 352 T; 0 other;
```

```
Qy     79  CAGGGCCGAGGGCCTCCGCCACTACTATTCTCAACATCCACGAGCATCAGCTCCTTCTT 138
              ||||| |      |||||  |       |   ||||||   | | | | || | |
Db     31  CATGAAGGTGGTATAAGACCATATTTTGAACACAACAAGATACCGAAAATAAGAAATA 90

Qy    139  CGCCAAAAGACTCATAACCTCAACCGTCTCGAGGCTCAGAGAAACGACCTCAATTCTAGG 198
```

```
Query Match           42.8%;  Score 536.6;  DB 24;  Length 1206;
Best Local Similarity 66.6%;  Pred No.  5.1e-156;
Matches  782;  Conservative  0;  Mismatches 389;  Indels  3;  Gaps  1;
```

Db   1168 GACGATGGTGCGTGCTCCTTACAGAAATTATTCAA 1202

Db   Aspergillus fumigatus essential gene #1557.

Query Match   40.7%; Score 510.4; DB 25; Length 1170;

Qy   91 CAGCAAAACACACAGAACTGCCGTCGTTGTGGAGCCCAAGGTAACAAATTGAATAACAA 150
              |||| ||| | | | | | |
Db   199 GTGAGGATGCTGGCGGCAAGAAATTACAGCTTCTGCGAGGAACCCGGCTCTTATGTGGTGGAA 258

Qy   151 GTCAGACAATTAAAGGATTGAAATTAATAATATTACAGAACCTGGATTCAATATGTAGGTTGAA 210
           ||||||| |||||||
Db   259 GTTGTCAAAGTAATGGCAAGAACAAAGTCCTTGTCACCCGAGACGAAAAATAT 318

Qy   211 GTTGTTAAAGTTATTGGGCTTGAAAAGAAGTTTTAGTGAAAAATCCATCCTTGAAGGTAAAATTC 270
           ||||||| ||||| ||| |||| |||| || ||||||
Db   319 GTTTGTTGACATTGACAATTGACATTCAAAGATTACTCCATCCACTAGAGTTGCA 378

Qy   271 AATTGTCAATGTTACCAAAGATATCCATCCTTCAGAAAATTGACTTCCATCAATAAGACTGTGT 330
           |||||| || || |||||||||| ||| |||| || ||
Db   379 CTCCGCAACGACGAGTTATGTTCTTCACTTAGTTCTGCCAAGTAAAGTTGATCCGGTC 438

Qy   331 TTGAAAACCAGATTCTCACGATTTGTTGTACAAAAATATTACCCACAACAAAGTTTGGTT 390
           ||||||| ||| | |||| ||| | |||| | |||
Db   439 AATCTGATGAAAAGTTGAGAAAGTTCCCGATTCTACATATGACATGAAATTGGTTGGTTTTGAC 498

Qy   391 TCATTGATGATGGTTGAAAAAAGTCCCAGAGCTCAACATACGATAATGGTTGGTGGGTTAGAT 450
           ||| | | ||| ||| ||| |||| | ||| | | || || |
Db   499 CAGCAAAATTAAAGAAAATAACAGGTGATTGAGCTACCAATCCATCCTGAGCCGTTT 558

Qy   451 AAACAAATTAAAGAAGTAATTAATTGGGGAAGAGGTTCTCGTAAGGTTGAACATTCCAGAATTGTTT 510
           |||| ||| || | | | || | | | || ||
Db   559 GAAAGTCCTTGGAATTGCAACAACCAAAAGGGTGTCCTCCGCCTGGTACAGGT 618

Qy   511 GAAAGTTTGGGTTATAGCTCACCAAAAGGTGTTTATATTGTAATGGCACACCAGGTTACAGGT 570
           |||| | ||| || ||| | |||||
Db   619 AAAACATTGTTGGCTAGGGCAGTGGGCTCATCATACTGACTGTACATTCATCAGGGTGTCT 678

Qy   571 AAGACCTTATTGGCAACAGACGCGTTGCTCACCAATACCGAATGTAAATTCATTCGTGTTTTCT 630
              ||| | | ||| || | |
Db   679 GGTTCTCGAGTTAGTTCAGAAATAACATTCGAGAATAGGTCAGGAACTTTT 738

Qy   631 GGTTCAGAATTAGTGCAAAAGTAATAATTGGGAGAGGGTTCGTATAGGTGGAAGGGTTCGGGAAGAGTGGTTC 690
Db   739 GTTATGGCCAGGAACAATGCCTCCATCAATTATCTTCATGGATGAAANTTGACAGTATTGGA 798

Qy   691 GTTGATGGCAAAGGAGAGCAATGCACCATCAATCATTTTTATGGATGAAAATTGATTCTATTGGG 750
Db   799 TCTGCTCGGAGTGGAAATCTGGAAGTGGGCAACGCGGTCAGTGTAGCCGTACTATCGCTG 858

Qy   751 TCGTCGGGTGTTGAAGGTTCATCAGG---AGGTGATTCCGAAGGTGCAAAGAACCAATGTTG 807
Db   859 GAACTTCTCAACCAGTTGGATGGCAGTTTCAAATAAGGATCAAGGTTTTGGATGGCC 918

Qy   808 GAATTGTTAAATCAATTGGATGGTTTTGAAAGTTCAAAAGAGTAGATATCAAAATCAATTATATGGCC 867
Db   919 ACCAATCGGATTGATATCCTGGATCAAGCCCTCCTTAGACCGAGGACCGGAAA 978

Qy   868 ACTAATCGTTTGAGATAGTTTGGATCCAGCTTTATTGAGACCAAGGTAGAATTGAATGAAAA 927
Db   979 ATTGAATTTCAAACCCCTTAAGATGAAGAGTTCTCGGGCTGCAGATAATTTTGAAAATCCATTCTAGA 1038

Qy   928 ATTGAATCCCAGCACCACCGTTGCTGCTGAGAACAGATATTTTAAAGATTCAATCAAGA 987
Db   1039 AGAATGAATTTAAATGCGTGGCCATTGATTGAAGAAGANTTGCCGAGAAGAGTGAATGGGAGCA 1098

Qy   988 TCCATGAAATTTGACAGAGGTTATATAAAATTGCTGAGAAAATGAAATGGCG 1047
Db   1099 TCTGGTGCTGAACTTAAGCTGTTTGCACTGAGACTGGAAGTTTTGCTTTTGAGGGGAGGCGG 1158

Qy   1048 AGTGGTGGCTGGAGTGGAAAGGAGTTGTACCGGAAGCTGGTATGTATGGCATTGAAGTTGAGAGAAGA 1107
Db   1159 AGGGTACACGTGACTGACTCAGGAGGAGATTTTGAGAATGGCCGTGGCGCAAGGTGATGAAAAGGAG 1218

Qy   1108 AGAATTCATGTAACTCAAGAAGATTTCGGAATGGCAGTTGCAAAAAGCTCATGTCTAAGGAAT 1167
Db   1219 ACTGAAAAAACACATGTCATTGCGGAAGTTGTGGGAA 1253

Qy   1219 ACTGAAAAAACACATGTCATTGCGGAAGTTGTGGGAA 1253

Best Local Similarity   65.4%;   Pred. No. 7.2e-148;
Matches 764;   Conservative   0;   Mismatches 401;   Indels   3;   Gaps   1;

```
Qy       87 GGGGCCTCCGCCACTACTATTCTCTCAACATCCACGAGCATCAGCTCCTTCTTCCGCCAAAA 146
            ||||| ||||||| | | | ||  ||  | ||  | | || | | ||  |  ||  ||| ||
Db        3 GGCGCTGGACAACTACTATCGCAATAAGATCGAGAGTATGAAACTCAGATTATCCAAGG 62

Qy      147 GACTCATAACCTCAACCGTCTCCGAGGGTTCAGAGAAACGAACCTCAATTCTAGGGTGAGGAT 206
            |||||  ||| || ||| ||  | ||| | ||||||| |  ||| || | ||||| | | |
Db       63 CCAGGGCTGTTTTACGTCGACTCGAAGCACAGCCCACGGACTAATAAATTCGCCAGTTCGCCT 122

Qy      207 GCTGCGCGAAGAATTACAACTTCTGCAGGAACCCGGCGCTCTTATGTCCGTGAAGTTGTCAA 266
            |||  | || | || ||| |  |||  || |||| || ||  ||||  ||| | | ||||
Db      123 ACTACGAGGAGAGTTGGGCTTGCTTCAACAACCTGGGGTCTTATGTGGGAGAGGTCGTTGAA 182

Qy      267 AGTAATGGGCAACAAACAAGTCCTTTGCCAAGGTCCACCCACGAAAAATATGTTGTTGA 326
            | |   || | |||| || ||  | |  | ||| |  | || |||||| | |||| |||
Db      183 GGTCATGAGCACCAAGAGAGGTTCTTGTCAAGGTACATCCAGAAGAGAAAATATGTCCGTTGA 242

Qy      327 CAATGACAAAAATATTGACATTACAAAGATTACTCCATCCACTAGACAGTTGCACTCCCCAA 386
            |||||| | ||| | |||  |||| |||| ||| || | || || || ||| |||| | |
Db      243 CAATGCAGACGGGGTCGATATTACGAAACTTACTGTGTGGGCAAGCGTGTTGGGTTGCTTTC 302

Qy      387 CGACAGTTATGTTCTTCACTTAGTTCTGCCAAGTAAAGTTGATCCATTGGTCAATCTGAT 446
            || | || | | |||   | || |  |||| || || ||| |||| |||||| ||| ||
Db      303 AGACTCTTATAAATTAGAAAAATAGTTGCCATGCTGCGTGTTGACCCGGTCGTTTCGCTTAT 362

Qy      447 GAAAGTTGAGAAAGTTCCCGATTCTACATATGACATGATTGGTTTGTTTAGACCAGCAAAT 506
            |  |||| | ||||  | || ||  ||   || ||| || ||| | ||||| ||| ||||
Db      363 GATGGTCGAGAAGGTGCCTGACAGCACATACGACAATATGATGGAGGCCTTGATCAACAAAT 422

Qy      507 TAAAGAAAATAAAGAGGTTCATTGAGCTACCAAATCAAACAATCCTGAGCTGTTTGAAAGTCT 566
            || |||| || || | || ||  | | ||  ||| || |  ||| |||| | ||||||||
Db      423 CAAGGAGATCAAGGAAGTCATTGAGGTGTTCTTTCGGCCGCACACCCGGTAAAAACACT 482

Qy      567 GTTGGCTAGGGCAGTGGCTCATCACTACTGACTGTACATTCATTCATCAGGGTGTTCTGA 626
            |||||| | |  |||  | | |||  | |||  | || |   | ||||||| |||| ||
Db      483 TGGTATCGGCACAGCCCAAAGGTGTTCTTCTTTACGGGCCGCAACCGGTAAAAACACT 542

Qy      627 GCTCGGCCCGAGCCGGTAGCCCATCACACAGATGCCGATTCACTCAGGGTCACGGGCTCGGA 686
            ||| |||  ||| || | |||||| |  | || || |||   |||||| | |||| ||||
Db      543 GTTCGGCTAGGGGCAGTGGCTCACATCACTACTGACTGTACATTCATTCATCAGGGTGGGTTCTGA 602

Qy      687 GTTAGTTCAGAAATACATTGGAGAAGTTCTAGAATGGTGGCGAACTTTTGTTATGGC 746
            ||| || || | ||  |||  | |||| ||| ||  | ||| | || ||||| | ||
Db      603 ACTAGTCCAGAATACATTGAGGGTAGTCGCATGGTGCTGTTCGTCATGGC 662

Qy      747 CAGGGAACATGCTCCATCAATTATCTTCATGGATGAAATTGACAGTATTGGATCTCGCTCG 806
            ||||||| ||||| || |||  | || || | |  || || | || | |||| | || ||
Db      663 TCGAGAACATGCACCAAGCATTATCTTCATTGACAGTGTTCGGTTCTGGAACTTCTTAGCCG 722

Qy      807 GATGGAATCTGGAAGTGGCAACCGTGATAGTGAGGTACAGCGTACTATGCTGGAACTTCT 866
            ||| || |  | | ||| | || | | | || | | |||  || | ||||||||  | ||
Db      723 TATAGACTCGG---CTGGCTCTGGAGATTCAGAGGGTGCGCACGTACAATGTGGAGTGCT 779

Qy      867 CAACCAGTTGGATGGATTGGAGAGCTTCAAATAAGATCGTTTTGATTGGTCACCAATCG 926
            || | || || | | | ||||| ||   || | | | || | || | |||| | |||
Db      780 CAATCAGCTGGATCGGTTTGAGCCCACAAAAAATCATCAACGGCTACGAACCG 839

Qy      927 GATGATANCCTGGATCAAGCCCTCCTTAGACCAGGACCGGATAGACCGGAAAATTGAAATT 986
            || | || || ||||||| | |  | || ||| | |||| | | || || || || ||||
Db      840 ACTTGATATCCTGGATCGGTTTGCCGGGCCGGGACGGAGGGATTGACCGGAGATCGAGTT 899

Qy      987 TCCAACCCCTAATGAAGAGTCTCGGCCTGGATATTTGAAAATCCATTCTAGAAGAATGAA 1046
            || | || |  || || | | ||  |  | ||| |||||| || ||| | |||| || ||
Db      900 TCCACCGCCATCCGTGGAAGCGTCGCCGATAATCTACGGCATTCACGTCGCATGAA 959

Qy     1047 TTTAATGCGTGGCATTGATTTGAAGAACAGATTGCCGAGAAGATGAATGGAGCAATCGGTGC 1106
            || | | || | ||| | ||| || || | || | || ||| || | || || | ||| ||
Db      960 CCTGACGCGGAGGCAATCAACACTTGACGGAGAAGATGGAAATGGAGGAATGGCTCAGGAGC 1019

Qy     1107 TGAACTTAAGGCTCGTTTGCACTGAAGCTGGAATGTTTGCTTGAGGGAGCGGAGGAGGTACA 1166
            |||||| || || || || |||| ||||| ||| | |  | ||||| | || ||| | ||
Db     1020 AGAACTGAAGGGTGTGGTGTCGACCGAGGGACGGCAGGCGCATGTACGGCGCGTCCA 1079
```

```
Db     1020 AGAACTGAAGGGTGTGTGTCGACCGAGGGACGGCAGGCGCATGTACGGCGCGTCCA 1079
Qy     1167 CGTGACTCAGGAGGATTTTGAGATGGCCCGTGGGCGAAGTGATGAAAAAGGAGACTGAAAA 1226
            ||| |  || |||||| || ||| |  | || || || || ||  ||| ||||  || ||
Db     1080 CGTCACGCAGGAGAAGATTTTGACCTAGCCACTGCCCAAGAATTCTCAACAAACACACGGATGACAA 1139

Qy     1227 AAACATGTCATTGCGGAAGTTGTTGGAAG 1254
            ||| || |  || || |||  |  |||||
Db     1140 GGAGGTTGCTTGTCTCTGGAAACTGTTCAAG 1167
```

Search completed: November   9, 2003, 09:03:39
Job time : 446 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2003, 08:54:25 ; Search time 104 Seconds
                                                (without alignments)
                                                5322.062 Million cell updates/sec

Title: US-09-462-972-1
Perfect score: 1254
Sequence: 1 atggcttcttgtaggagttga.............cattgcggaagttgtggaag 1254

Scoring table: IDENTITY_NUC
                     Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_NA:*
                     1: /cgn2_6/ptcdata/1/ina/5A_COMB.seq:*
                     2: /cgn2_6/ptcdata/1/ina/5B_COMB.seq:*
                     3: /cgn2_6/ptcdata/1/ina/6A_COMB.seq:*
                     4: /cgn2_6/ptcdata/1/ina/6B_COMB.seq:*
                     5: /cgn2_6/ptcdata/1/ina/pCTUS_COMB.seq:*
                     6: /cgn2_6/ptcdata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	Match DB	DB ID	Description
1	312	24.9	1664976	4	US-08-916-421B-1	Sequence 1, Appli
2	270.6	21.6	1167	2	US-08-820-170A-14	Sequence 14, Appl
3	270.6	21.6	1167	2	US-09-055-699-14	Sequence 14, Appl
4	270.6	21.6	1167	3	US-09-273-565-14	Sequence 14, Appl
5	270.6	21.6	1167	4	US-09-565-538-14	Sequence 14, Appl
6	270.6	21.6	1167	4	US-09-661-468-14	Sequence 14, Appl
7	270.6	21.6	1167	4	US-09-976-165-14	Sequence 14, Appl
8	270.6	21.6	1566	2	US-08-820-170A-15	Sequence 15, Appl
9	270.6	21.6	1566	3	US-09-055-699-15	Sequence 15, Appl
10	270.6	21.6	1566	3	US-09-273-565-15	Sequence 15, Appl
11	270.6	21.6	1566	4	US-09-565-538-15	Sequence 15, Appl
12	270.6	21.6	1566	4	US-09-661-468-15	Sequence 15, Appl
13	270.6	21.6	1566	4	US-09-976-165-15	Sequence 15, Appl
14	203.2	16.2	727	4	US-08-998-416-600	Sequence 600, App
15	192.2	15.3	277	4	US-09-313-294A-1050	Sequence 1050, App
16	189.2	15.1	263	4	US-09-313-294A-701	Sequence 701, App
17	173.6	13.8	811	4	US-08-998-416-414	Sequence 414, App
18	168.8	13.5	261	4	US-09-313-294A-543	Sequence 543, App
19	163.4	13.0	1230025	4	US-09-198-452A-1	Sequence 1, Appli
20	161.8	12.9	21338	4	US-08-961-527-20	Sequence 20, Appli
21	157	12.5	1959	3	US-08-987-123-1	Sequence 1, Appli
22	157	12.5	1959	3	US-09-987-123-1	Sequence 1, Appli
23	151.4	12.1	5919	4	US-09-328-352-1369	Sequence 1369, Ap
24	148	11.8	1911	4	US-09-987-123-4	Sequence 4, Appli
25	144.2	11.5	2106	3	US-09-134-001C-490	Sequence 490, App
26	142	11.3	2932	3	US-08-999-774A-5	Sequence 5, Appli
27	142	11.3	2933	4	US-09-149-476-165	Sequence 165, App
28	142	11.3	3276	4	US-09-149-476-298	Sequence 298, App
29	140.8	11.2	580073	4	US-08-545-528D-1	Sequence 1, Appli
30	138	11.0	2094	4	US-08-816-177-1	Sequence 1, Appli
31	135.2	10.8	2417	4	US-09-216-393B-306	Sequence 306, App
32	135.2	10.8	2417	4	US-09-216-393B-308	Sequence 308, App
33	128	10.2	1830121	4	US-09-557-884-1	Sequence 1, Appli
34	128	10.2	1830121	4	US-09-643-990A-1	Sequence 1, Appli
35	127.8	10.2	2488	2	US-09-816-093-1	Sequence 1, Appli
36	122.4	9.8	709	3	US-09-328-111-794	Sequence 794, App
37	118.2	9.4	7060	4	US-09-221-017B-479	Sequence 479, App
38	117.8	9.4	1227	4	US-09-107-532A-2599	Sequence 2599, Ap
39	112.4	9.0	4403765	2	US-09-103-840A-2	Sequence 2, App
40	112.4	9.0	4411529	3	US-09-103-840A-1	Sequence 1, App
41	98.4	7.8	2837	2	US-09-313-294A-1930	Sequence 1930, Ap
42	98.4	7.8	265	4	US-08-993-228-11	Sequence 11, Appl
43	97.6	7.8	2837	4	US-09-134-001C-476	Sequence 476, App
44	96.2	7.7	296	4	US-09-313-294A-4379	Sequence 4379, Ap
45	95.4	7.6	817	3	US-08-998-416-372	Sequence 372, App

ALIGNMENTS

RESULT 1
US-08-916-421B-1

Sequence 1, Application US/08916421B
Patent No. 6503729
Patent No. 6503729
GENERAL INFORMATION:
APPLICANT: Bult et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoc
TITLE OF INVENTION: Jannaschii
FILE REFERENCE: PB275
CURRENT APPLICATION NUMBER: US/08/916,421B
CURRENT FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
 LENGTH: 1664976
 TYPE: DNA
 ORGANISM: Methanococcus jannaschii
 FEATURE:
   NAME/KEY: misc_feature
   LOCATION: (28222)..(28222)
   OTHER INFORMATION: n equals a, t, c, or g
   NAME/KEY: misc_feature
   LOCATION: (2825)..(28258)
   OTHER INFORMATION: n equals a, t, c, or g
   NAME/KEY: misc_feature
   LOCATION: (84773)..(84773)
   OTHER INFORMATION: n equals a, t, c, or g
   NAME/KEY: misc_feature
   LOCATION: (84808)..(84808)
   OTHER INFORMATION: n equals a, t, c, or g
   NAME/KEY: misc_feature
   LOCATION: (84812)..(84812)
   OTHER INFORMATION: n equals a, t, c, or g
   NAME/KEY: misc_feature
   LOCATION: (98120)..(98120)
   OTHER INFORMATION: n equals a, t, c, or g
   NAME/KEY: misc_feature
   LOCATION: (98159)..(98159)
   OTHER INFORMATION: n equals a, t, c, or g
   NAME/KEY: misc_feature
   LOCATION: (98239)..(98239)
   OTHER INFORMATION: n equals a, t, c, or g
   NAME/KEY: misc_feature
   LOCATION: (98266)..(98266)
   OTHER INFORMATION: n equals a, t, c, or g
   NAME/KEY: misc_feature
   LOCATION: (98343)..(98343)

OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1039908)..(1039908)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (242220)..(242220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (600592)..(600592)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1664855)..(1664855)
OTHER INFORMATION: n equals a, t, c, or g

US-08-916-421B-1

Query Match 24.9%; Score 312; DB 4; Length 1664976;
Best Local Similarity 57.2%; Pred. No. 2.5e-85;
Matches 586; Conservative 0; Mismatches 435; Indels 3; Gaps 1;

```
Qy      204 GATGCTGCGCGAAGAATTACAGCTTCTGCAGGAACCGGGCTCTTATGTCGGTGAAGTTGT 263
                ||||||||||  ||||||    |||    ||
Db   356314 GATTTTAAGAAGAGAATTAGATAGAAATGAGAGTCCCTCCATTGATTAGTTGGAACTGTAGT 356373

Qy      264 CAAAGTAATGGGCAAGAAACAAAGTCCTTGTCAAGGTCACCCGAGAAGGAGAAAATATGTTGT 323
                ||||  |||||| ||||
Db   356374 TGATAAAGTAGGAGAGAGGAAGAAAAGTAGTTGTTCAAAAGCTCAACAGGCCCAAGTTTCTTAGT 356433

Qy      324 TGACATTGACAAAAATATTACATTACAAAGATTACTCCATCCACTAGAGTTGCACTCCG 383
                | |||
Db   356434 TAATGTCTCTCACTTTGTAAATCCAGATGAGTTTAGCCCCTGGGAAAGAGAGAGTCTGTTTAAA 356493

Qy      384 CAACGACAGTTATGTTGTTCTTCACTTAGTTCTTGCCAAGTAAAAGTTGATCCATTGGTCAATCT 443
```

```
Db  356494  TCAGCAAACATTAACAGTTCTGTTGATGTATTGCCAGAAAATAAAGACTACAGAGGCTAAAGC  356553
                  |||              |||                                 |||
Qy     444  GATGAAAGTTGAGAAAGTTCCCGATTCACATATGACATGATTGGTTGGGTTTAGACCAGCA      503

Db  356554  AATGGAAGTTGATGAAAGACCAAATGTTAGCAATGATTAGCATATGAGATATTGGTGGAAAACA  356613
                 |||||                |||     |||     |||  |||
Qy     504  AATTAAGAAATAAAAGAGGTCATCATGTGAGCTACCAATCAAACATCCTGAGCTGTTTGAAAG    563

Db  356614  AATGCAAGAAAATTAGAGAAGTTGTTGAACTCCCATTGAAACATCCAGAATTGTTTGAAAA  356673
                    |||  |||            |||              |||  |||
Qy     564  TCTTGGAATTGCACAACCAAAGGGTGTCCTGCTCTATGGGCCACCTGGTACACGGTAAAAC    623

Db  356674  GGTTGGAATTGAACCACCAAAAGGTATTCTGCTTTACGGACCACCAGGAACTGGAAAGAC  356733
                  |||   |||   |||        |||       |||          |||
Qy     624  ATTGTTGGCTAGGGCAGTGGCTCATCATCACTGTCATTCATCGGGTGTCTCGGTTC      683

Db  356734  ATTATTAGCTAAAGCTGTTTGCTACAGAACAAAATGCTACCTTTATAAGAGTTGTTGGGTTC  356793
                 |||                |||    |||           |||
Qy     684  TGAGTTAGTTCAGAAATACAATGGAGGAAGGTTCTAGAATGGGTCAGGGAACTTTTTGTTAT    743

Db  356794  TGAATTGGTTAAGAACTTTTATTGGAGGGGGGCTTCGTTAGTTAAAGATATATTCAAATT  356853
                   |||          |||                    |||  |||
Qy     744  GGGCCAGGGAACATGCTCCATCAATTATCTTCACGTAACAGTAATTGCACGTAAAAC      803

Db  356854  GGCTAAAGAAAAAAGCTCCTCAATCATATTCATATAGATGACATTGATTGCTATTGCAGCAAA  356913
                    |||          |||          |||       |||
Qy     804  TCGGGATGGAATCTGGAAGTGGCCAACGGTGATTAGTGAGGTGTACACGGTACTTATGCTGGAACT    863

Db  356914  GAGAACAGACGCTTTAACTGGT---GGAGGATAGGGAAGTTCAGAGAACATTAATGCAGTT  356970
                 |||              |||            |||      |||
Qy     864  TCTCAACCAGTTGATGGATTGAAGCTTCAAATAGAGATCAAGGTTTGATGGCCACCAA      923

Db  356971  GTTGGCAGGATGGATCGATCGAACGGAGCAGTTAAACAGTTAACTGAAGAAAAATCATGTGTTAAAGCTAACCAATGGA  357030
                  |||             |||          |||      |||
Qy     924  TCGGATTGATATCCTGGATCAAGCCCTCTTAGACCAGCAGACGGAAAATTGA      983

Db  357031  CAGACCTGACATTTTAGACCCTGCAATATTAAGGACCTGAAGAATTGGATAGAATCATAGA  357090
                   |||      |||          |||      |||
Qy     984  ATTTCCAACCCTTAATCAAGAGACTCTCGGCTCTGGATATTTTTGAAAATCCATTCTAGAAGAAT      1043

Db  357091  AGTCCCAGCTCCTTGATCGGAGCGGGGTGAGTATTCATCAACTACAAGAAAAAAGAT  357150
                   |||      |||          |||      |||
Qy    1044  GAATTTAAATGCGTGGCCATTGATTTGAAAGAGATGCCGGAAGAAGATGAAATGGGCATCTGG      1103

Db  357151  GAATTTAGCCGGAAGATGTCAATTTAGAGAAATAGCTAAGACTGAAGGATGTGTAGG  357210
                   |||      |||      |||
Qy    1104  TGCTGAACTTAAGGCTCTTTCCACTGAAGCTGGAATGTTTGCTTTGACGGGAGCGGGAGGT      1163

Db  357211  GGCTGAGTTAAAGGCAATCTCCACAGCAGGAGGGATGAATGCAATTAGGAGTTAAGGGA  357270
                   |||      |||      |||
Qy    1164  ACACGTGACTCAGGGAGGGGATTTGAGATGGCCGGTGGCCGAAGGTGATTGAAAAAAGAGAGACTGA      1223

Db  357271  CTATGTAACAATGGATGGAACACTTTAGAAGGTTGAGAAGATTATTGGGACAAGAAAAGT  357330
                   |||      |||
Qy    1224  AAAA      1227

Db  357331  TAAA  357334
```

```
Query Match          21.6%;   Score 270.6;  DB 2;   Length 1167;
Best Local Similarity 55.9%;  Pred. No. 2.4e-74;
Matches 535; Conservative 0; Mismatches 419; Indels 3; Gaps 1;

Qy     250  GTCGGTGAAGTTGTCAAAGTAATGGGCAAGAACAAAGTCCTTGTCGTCCACCCAGAA      309
                  |||      |||      |||
Db     169  GTGGGTGAAGTGCTTAAACAGTTAAACTGAAGAAAAATCATGTGTTAAAGCTAACCAATGGA  228

Qy     310  GGAAAATATGTTGTTGACAATGACAAAAATAATTGACATTACAAAAGAATTACTCCATCACT      369
                  |||      |||      |||
Db     229  CCAAGATATGTTGTGGGTTGTCCGACAGCTTGACAGAAAAGTAAGCCTGAAGCCAGGAAACA  288

Qy     370  AGAGTTGCACTCCGCAACCACACGTTATGTTCTTCACTTAGTTCTGCCAGTAAAGTTGAT      429
                  |||      |||      |||
Db     289  AGAGTTGCTTTGGATATAGACTACACTAACTATCATGAGATATTGCCCGAGAGAGGTGGAT  348

Qy     430  CCATTGGTCAATCTGATGAAAGTTGAGAAAGTTCCCAATTCTACATATGCATTGATTGGT      489
                  |||      |||      |||
Db     349  CCACTCGGTTTATAACATGTCTCATGAGGACCCTGGGAATGTTCTTATTCTGAGATTGGA  408

Qy     490  GGTTTAGACCAGCAAAATTAAAGAAAATAAAGAGGTCATTGAGGTACCAATCAAACATCCT      549
                  |||      |||      |||
Db     409  GGGCTATCACGAACAGATCCGGGAATTAAGAGAGGGTGATAGAAGTTAAGGCTGAGAACCA  468

Qy     550  GAGCTGTTTGAAAGTCTTGGAAATGCACAACCAAAGGGTGTCCTGCTCTATGGGCCACCT      609
                  |||      |||      |||
Db     469  GAGTTATTTCAGCGTGTAGGAATTAATACCTCCAAAAGGCTGTTTGTTATATGGGACCACCA  528

Qy     610  AGGGTGTCTGGTTCTCGAGTTAGTTCAGAAATACATTGGAGAAGGTTCTAGAATGGTCAGG      669
                  |||      |||      |||
Db     529  GGTACGGGAAAAACACTCTTGGCACGAGCCGGTTGCTAGCGACTGCTGGCCAATTCTTA  588

Qy     670  AGGGTGTCTGGTTCTCGAGTTAGTTCAGAAATACATTGGAGAAGGTTCTAGAATGGTCAGG      729
                  |||      |||      |||
Db     589  AAGGTGTATCTAGTTCTTATTGTAGACAAGTACATTGGTGATGAATGCTCGGTTTGATCGA  648

Qy     730  GAACTTTTTGCTTATGGGCCAGGGAACATCCTCCATCATCAATTATCTTCATCGGATGAAATTGAC      789
                  |||      |||      |||
Db     649  GAAATGTTTAATTATGCTAGAGATCAATCAAACCAATGCATCATTTTTATAGGATGAAAATAGAT  708

Qy     790  AGTATTGGATCTGCTCGGATGGAATCTCTGGAAGTGGCAACGGGTGAATAGTAGGTACACGGT      849
                  |||      |||      |||
Db     709  GCTTATGGGTGGTCGGTCGGTTT---TCTGAGGGGTTACTTCAGCTGACACAGAGAGATTCAGAGA  765

Qy     850  ACTAGCTGAAACTTCTCAACCAGTTGGATGGAATTTGAAAGCTTCAAATAAGATCAAGGTT      909
                  |||      |||      |||
```

Db 766 ACGTTAATGGAGTTACTGAATCAAATGGATGGATTTGATACTTCTGATAGAGTTAAAATG 825

Qy 910 TTGATGGCCACCAATCCGGATTGATATCCTGATCAAGCCTCCTTAGACCAGGACGGATA 969
Db 826 ACCATGGCTACAAACAGACAACCAGATACACTGGAATCCTCGCTTGCTGCTGCTCCAGGAAGATTA 885

Qy 970 GACCGGAAAATGAAATTCCAACCCTTAATGAGAGTCTCGGCTGGATATTTTGAAAATC 1029
Db 886 GATAGAAAAATACATATTGATTGGCCAAATGAAGAACAGCCAAGATTAGACATACTGAAAATC 945

Qy 1030 CATTCTAGAAGAAATGAATTAATGCCTGGCATTGATTTTGAAGAAGATTGCCGAGAAGATG 1089
Db 946 CATGCAGGTCCCATTACAAAGCCATGGTGAAAAATGATATTATGAAGCAATTGTGAAGCTTTCG 1005

Qy 1090 AATGGAGCATCTGGTGCTGAACTTAAGGCTCGTTTGCCACTGAAGTGTTGCTTTG 1149
Db 1006 GATGGCTTTAATGGAGCAGATCTGAGAAAATGTTTGTTACTGAAGCAATTGTTCGGCAATT 1065

Qy 1150 AGGGAGCGGAGGGTACACGTGACTCAGGAGGAGGAATTTGAGATGGGACCGTGGGCCAAGGTG 1206
Db 1066 CGTGCTGATCATGATTTTGTAGTACAGGAAGAGACTTCATGAAAGCAGTCAGAAAAGTG 1122

US-09-055-699-14
Query Match   21.6%;   Score 270.6; Pred. No. 2.4e-74; DB 3; Length 1167;
Best Local Similarity   55.9%;
Matches 535; Conservative 0; Mismatches 419; Indels 3; Gaps 1;

Qy 250 GTCGGTGAAGTTGTCAAAGTAATGGGCGAAGAGACAAAAGTCCTTGTCTCAAGGTCCACCCAGAA 309
Db 169 GTGGGTGAAGTGCTAAACAGTTAACTGAAGAAAAATTCATTCGTTGTTAAAGCTACCAAATGGA 228

RESULT 4
US-09-273-565-14

Qy 310 GGAAAATATGGTTGTTGACATTGACAAAAAATATTGACATTACAAAAGATTACTCCATCCAT 369
Db 229 CCAAGATATGTTGTTGGGTTGTCGTCGACACGTTGACAAGAGCTTGACAAAAGTAAGGCAGGAAACA 288

Qy 370 AGAGTTGCACTCCGCAACAGCACAGTTAATGTTCTTCACTTAGTTCTGCCAAGTAAAGTTGAT 429
Db 289 AGAGTTGCTTTGGGATATGATGACTACACTAACTATCACATGAGATAGATAGGTGGAT 348

Qy 430 CCATTGGTCAATCTGATGAAAGTTGAGAAAGTTCCCGGATTCTACATATGACATGATTGGT 489
Db 349 CCACTGGTTTATAAACAATGTCTCATGAGGACCCCTGGAAATGTTTCTTATTCTGAGATTGGA 408

Qy 490 GGTTTAGACCAGCAAATTAAAGAAATAAAAGAGGTCATTGAGCTACCAAATACAAA 549
Db 409 GGGCTATCAGAACAGATCCGGGAATTTAAGAGGTGAATAGAATTAACCTCTTCACAACCA 468

Qy 550 GAGCTCGTTGAAAGTCTTGGAAATTGCACAACCAAAGGGTGTCCTGCTCTATGGGCCACT 609
Db 469 GAGTTATTCACCGCTGTAGGGAATAATACCTCCAAAAGGCTGTTTTGTTATATTGGACACCA 528

Qy 610 GGTACAGGTAAAACATTGTTGGGCTAGGCGGAGTGGGGCTCATCATACTGACGTACATTCATC 669
Db 529 GGTACGGGAAAAACACTCTTGGCACGAGCCGTTGCTTAGACCAGCTGGACTGCAATTCTTA 588

Qy 670 AGGGTGTTCTCGGTTTCGAGTTAGTTCAATACATTGGAGAAGGTTCTAGAATGGTTCAGG 729
Db 589 AAGGTTGTATCTAGTTCTTATTGTAGACAAGTACACTGGGAAAAGTGCTCGTTTGATCAGA 648

Qy 730 GAAACTTTTGTTATGGCCCACGGGAACATGGTCCATCAAATTTATCTTCATGGATGAAATTGAC 789
Db 649 GAAAATGTTTAATTATGCTAGAGATCATCACCATGGCATCATTTATTGGGATGAAATAGAT 708

Qy 790 AGTATTGGATCTGCTCCGGATGGGAATCTGGAAGTGGCAACGGTGATAGTGAGGTACAGCGT 849
Db 709 GCTATTGCGGTTGGTCGTCCGGTTT---TCTGAGGGTACTTCAGCTGACAGAGGAATTCAGAGA 765

Qy 850 ACTATGCTGGAACTTCTCAACCCAGTTGGATTGGGAGTTGAAAATAAGACTCAAGGTT 909
Db 766 ACGTTAATGGAGTTACTGAATCAAATGGATGGATTTGATACTTCTGATAGAGTTAAAATG 825

Qy 910 TTGATGGCCACCAATCCGGATTGATATCCTGATCAAGCCTCCTTAGACCAGGACGGATA 969
Db 826 ACCATGGCTACAAACAGACAACCAGATACACTGGAATCCTCGCTTGCTGCTGCTCCAGGAAGATTA 885

Qy 970 GACCGGAAAATGAAATTCCAACCCTTAATGAGAGTCTCGGCTGGATATTTTGAAAATC 1029
Db 886 GATAGAAAAATACATATTGATTGGCCAAATGAAGAACAGCCAAGATTAGACATACTGAAAATC 945

Qy 1030 CATTCTAGAAGAAATGAATTAATGCCTGGCATTGATTTTGAAGAAGATTGCCGAGAAGATG 1089
Db 946 CATGCAGGTCCCATTACAAAGCCATGGTGAAAAATGATATTATGAAGCAATTGTGAAGCTTTCG 1005

Qy 1090 AATGGAGCATCTGGTGCTGAACTTAAGGCTCGTTTGCCACTGAAGTGTTGCTTTG 1149
Db 1006 GATGGCTTTAATGGAGCAGATCTGAGAAAATGTTTGTTACTGAAGCAATTGTTCGGCAATT 1065

Qy 1150 AGGGAGCGGAGGGTACACGTGACTCAGGAGGAGGAATTTGAGATGGGACCGTGGGCCAAGGTG 1206
Db 1066 CGTGCTGATCATGATTTTGTAGTACAGGAAGAGACTTCATGAAAGCAGTCAGAAAAGTG 1122

; CURRENT APPLICATION NUMBER: US/09/273,565A
; CURRENT FILING DATE: 1999-03-22
; EARLIER APPLICATION NUMBER: 09/055,699
; EARLIER FILING DATE: 1998-04-07
; EARLIER APPLICATION NUMBER: 08/820,170
; EARLIER FILING DATE: 1997-03-19
; EARLIER APPLICATION NUMBER: 1997-03-19
; EARLIER APPLICATION NUMBER: JP 63410/1996
; EARLIER FILING DATE: 1996-03-19
; EARLIER APPLICATION NUMBER: JP 69163/1997
; EARLIER FILING DATE: 1997-03-05
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 14
; LENGTH: 1167
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-273-565-14

Query Match   21.6%;   Score 270.6; DB 3; Length 1167;
Best Local Similarity   55.9%; Pred. No. 2.4e-74;
Matches 535; Conservative 0; Mismatches 419; Indels 3; Gaps 1;

```
Qy   250 GTCGGTGAAGTTGTCAAAGTAATGGGCAAGAACAAAAGTCCTTGTCAAGGTCCACCCAGAA 309
             ||||| |||   |||  ||||  ||||  ||   ||||   |||||| |||||||||||
Db   169 GTGGGGTGAAGTGGCTTAAACAGTTAACTGAAGAAAAATTCATTGTTAAAGCTACCAATGGA 228

Qy   310 GGAAAATATGTTGTTGACATTGACAAAAATATTGACATTACAAAAGATTACTCCACTCCACT 369
             |||||| || |||||||   ||  |||||||||    ||| |||||   || || ||||||
Db   229 CCAAGATATGTTGTTGGGGTGTTCGTCGACAGCTTGACAAAAGTAAGCTGAAGCCAGGAACA 288

Qy   370 AGAGTTGCACTCCGCAACGACAGTTATATGTTCTTCACTTAGTTCTGCCAAGTAAAGTTGAT 429
             |||||||||| |||| |||   || ||||| |||| |||||   || |||||| |||||||
Db   289 AGAGTTGCTTGGATATGACTAACTACACTAAATCATGAGAGTATATTCCGAGAGAGTGGGAT 348

Qy   430 CCAATTGGTCAATCTCGATGAAAAGTTGAGAAAAGTTCCCGGATTCTTACATAATGACGATTGGGT 489
             ||||||   |||| ||| || |  |||||  ||  ||  |  |||||| ||| || |||||| |||
Db   349 CCACTCGGTTTATAACATGTCTCATGAGGACCCTGGGAAGTGTTTCTTAATTCTGAGAATTGGA 408

Qy   490 GGTTTAGACCAGCAAAATTAAAGAAAATAAAAAGAGGTCATTCAGACTACCAATCAAAACATCCT 549
             ||| ||| ||||| |||||  |||| ||| |||| ||| |||  |||| ||   ||| |||||||
Db   409 GGGCTATCAGAACAGATCCGGGAATTAAGGAGAGGTTGATAGAGAATTACCTCTTAATTCGGACCACCA 468

Qy   550 GAGCTGTTTGAAAGTCTTGGAATTGCACAACCAAAAGGGTGTTCCTCGTTCTTATATGGGGCCACCT 609
             |||||| |||| ||| || || ||  |||| ||   |||||  ||  |||| |||| || ||||||
Db   469 GAGTTATTTCAGCGTGTAGGAGAATAATACCTCCCAAAAGGCTGTTTCTTTATATGGACCACCA 528

Qy   610 GGTACAGGGTAAAAACAATTGTTGGGCTAGGCGGAGTGGCTCATCACTATACTGACAGTCAATTCATC 669
             ||| ||||| || |||| || ||||| ||   ||  || || ||||  ||| ||||| |||||||||
Db   529 GGTACCGGGAAAAAACACTCTTGCTACGCCGTTGCTACCGGACTCGAACTGCCAATTTCTTA 588

Qy   670 AGGGTGTTCTCGGTTCTGAGTTAGTTCAGAAAATACATTGGAGAAGGTTCTGAATGGTCAGG 729
             |||||||| ||||||| ||||| || ||||  |||| |||||||||||| |||||||| ||
Db   589 AAGGGTGTATCCTAGTTCTATTGTGAGAAGTGCTCCATCAATTACTTCATGAGTGAAATTGAC 648

Qy   730 GAACTTTTGTTATGCCAGGGAACATGCTCCATCAATTATTCTTCATGAATGAAATTGAC 789
             ||||   || ||| ||||  |||||| || |||||| |||| || ||||||||||||||
Db   649 GAAAGTTTAATTATTGCTAGAGAATCATCAACCATGCAATCATTTTTATGGGATGAAAATAGAT 708

Qy   790 AGTAATTGGGATCTGCTGGATGGAATCTGAAAGTTGCAAGCGGATGGAGTACGAGCGCGT 849
             ||||  |||||| || |||    ||| ||  |||||| |||   |||  |||| | | |||
Db   709 GCTATTGGTGGTGCGTCGGTTT---TCTGAGGGTGACTTCAACTGACAGAGAATTCAGAGA 765

Qy   850 ACTATGCTGGAACTTCTCAACCAGTTGGATGGAGTTGAAGCTTCAAAATAAGATCAAAGGTT 909
             ||| || ||||| ||| || ||||||||||||||| ||||| ||||  |||||| |||||||
Db   766 ACGTTAATGGAGTTACTGAATCAAATGGATGGATTTGAAGCTTCAACTGACAGAGAGTTAAAAATG 825

Qy   910 TTGATGGCCACCAATCGGATTGATATCCTGGATCAAGCCCTCCTTAGACCAGGAGGACGGATA 969
             ||| || || |||||  ||||| || |||  |||||  || |||| ||  ||||| |||||
Db   826 ACCATGGCTACAACAAGACCAGATACCACTGGAATCTCGCTTTGCGTCGGTCCAGGAAGAGGATTA 885

Qy   970 GACCGGAAAAATTGAATTTCCAACCCCTATGACTGGAATATTTTGAAAATC 1029
             |||| ||||||  ||||  || |||   ||| ||||  |||||||||||||
Db   886 GATAGAAAATAACTATTATTGCCAATGAACAAGCAAGATTAGACTACTGAAAAATC 945
```

RESULT 5

US-09-565-538-14

; Sequence 14, Application US/0956553B
; Patent No. 6333404
; GENERAL INFORMATION:
; APPLICANT: FUJIWARA, TSUTOMU
; APPLICANT: WATANABE, TAKESHI
; APPLICANT: HORIE, MASATO
; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
; TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
; FILE REFERENCE: Q-53599
; CURRENT APPLICATION NUMBER: US/09/565,538
; CURRENT FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US/09/273,565
; PRIOR FILING DATE: 1999-03-22
; PRIOR APPLICATION NUMBER: 09/055,699
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: 08/820,170
; PRIOR FILING DATE: 1997-03-19
; PRIOR APPLICATION NUMBER: JP 63410/1996
; PRIOR FILING DATE: 1996-03-19
; PRIOR APPLICATION NUMBER: JP 69163/1997
; PRIOR FILING DATE: 1997-03-05
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 14
; LENGTH: 1167
; TYPE: DNA
; ORGANISM: Homo sapiens

Query Match   21.6%;   Score 270.6; DB 4; Length 1167;
Best Local Similarity   55.9%; Pred. No. 2.4e-74;
Matches 535; Conservative 0; Mismatches 419; Indels 3; Gaps 1;

```
Qy   250 GTCGGTGAAGTTGTCAAAGTAATGGGCAAGAACAAAAGTCCTTGTCAAGGTCCACCCAGAA 309
             ||||| |||   |||  ||||  ||||  ||   ||||   |||||| |||||||||||
Db   169 GTGGGGTGAAGTGGCTTAAACAGTTAACTGAAGAAAAATTCATTGTTAAAGCTACCAATGGA 228

Qy   310 GGAAAATATGTTGTTGACATTGACAAAAATATTGACATTACAAAAGATTACTCCACTCCACT 369
             |||||| || |||||||   ||  |||||||||    ||| |||||   || || ||||||
Db   229 CCAAGATATGTTGTTGGGGTGTTCGTCGACAGCTTGACAAAAGTAAGCTGAAGCCAGGAACA 288

Qy   370 AGAGTTGCACTCCGCAACGACAGTTATATGTTCTTCACTTAGTTCTGCCAAGTAAAGTTGAT 429
             |||||||||| |||| |||   || ||||| |||| |||||   || |||||| |||||||
Db   289 AGAGTTGCTTGGATATGACTAACTACACTAAATCATGAGAGTATATTCCGAGAGAGTGGGAT 348

Qy   430 CCAATTGGTCAATCTCGATGAAAAGTTGAGAAAAGTTCCCGGATTCTTACATAATGACGATTGGGT 489
             ||||||   |||| ||| || |  |||||  ||  ||  |  |||||| ||| || |||||| |||
Db   349 CCACTCGGTTTATAACATGTCTCATGAGGACCCTGGGAAGTGTTTCTTAATTCTGAGAATTGGA 408

Qy   490 GGTTTAGACCAGCAAAATTAAAGAAAATAAAAAGAGGTCATTCAGACTACCAATCAAAACATCCT 549
             ||| ||| ||||| |||||  |||| ||| |||| ||| |||  |||| ||   ||| |||||||
Db   409 GGGCTATCAGAACAGATCCGGGAATTAAGGAGAGGTTGATAGAGAATTACCTCTTAATTCGGACCACCA 468

Qy   550 GAGCTGTTTGAAAGTCTTGGAATTGCACAACCAAAAGGGTGTTCCTCGTTCTTATATGGGGCCACCT 609
             |||||| |||| ||| || || ||  |||| ||   |||||  ||  |||| |||| || ||||||
Db   469 GAGTTATTTCAGCGTGTAGGAGAATAATACCTCCCAAAAGGCTGTTTCTTTATATGGACCACCA 528
```

RESULT 6
US-09-661-468-14
Sequence 14, Application US/09661468
Patent No. 6376189
GENERAL INFORMATION:
Patent No. 6376189
APPLICANT: FUJIWARA, TSUTOMU
APPLICANT: WATANABE, TAKESHI
APPLICANT: HORIE, MASATO
TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
FILE REFERENCE: Q-53599
CURRENT APPLICATION NUMBER: US/09/661,468
CURRENT FILING DATE: 2000-09-13
PRIOR APPLICATION NUMBER: 09/055,699
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: 08/820,170
PRIOR FILING DATE: 1997-03-19
PRIOR APPLICATION NUMBER: JP 63410/1996
PRIOR FILING DATE: 1996-03-19
PRIOR APPLICATION NUMBER: JP 69163/1997
PRIOR FILING DATE: 1997-03-05
NUMBER OF SEQ ID NOS: 95
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 14
LENGTH: 1167
TYPE: DNA
ORGANISM: Homo sapiens
US-09-661-468-14

Query Match   21.6%;
Best Local Similarity 55.9%;   Score 270.6; DB 4;
Matches 535; Conservative 0; Mismatches 419; Length 1167;
                                                      Pred. No. 2.4e-74;
                                                      Indels 3; Gaps 1;

RESULT 7
US-09-976-165-14
Sequence 14, Application US/09976165
Patent No. 6562947
GENERAL INFORMATION:
APPLICANT: FUJIWARA, TSUTOMU
APPLICANT: WATANABE, TAKESHI

[DNA sequence alignment blocks — Qy (query) and Db (database) paired rows with position numbers, for both RESULT 6 (left column) and RESULT 7 (right column)]

; APPLICANT: HORIE, MASATO
; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
; TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
; FILE REFERENCE: Q-53599
; CURRENT APPLICATION NUMBER: US/09/976,165
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 09/565,538
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: 09/055,699
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: 08/820,170
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: 1997-03-19
; PRIOR FILING DATE: 1997-03-19
; PRIOR APPLICATION NUMBER: JP 63410/1996
; PRIOR FILING DATE: 1996-03-19
; PRIOR APPLICATION NUMBER: JP 69163/1997
; PRIOR FILING DATE: 1997-03-05
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 14
; LENGTH: 1167
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-976-165-14

Query Match   21.6%;   Score 270.6; DB 4; Length 1167;
Best Local Similarity   55.9%;   Pred. No. 2.4e-74;
Matches   535; Conservative   0; Mismatches   419; Indels   3; Gaps   1;

Qy   250 GTCGGTGAAGTTGTCAAAGTAATGGGCAAGAACAAAGTCTCCTTGTCCACTGACTCCACCAGA 309
           |||||| |||||||||| |||||| ||| |||||||||||||| ||||||||||| |||||||
Db   169 GTGGGTGAAGTGCTTAAACACAGTTAACTGAAGAGAAAAATTCATTGTTAAAGCTACCAATGGA 228

Qy   310 GGAAAATAGTTGTTGACAATGACAAAATATTGACATTACAAAGATTACTCCATCCACT 369
           |||| ||||||| ||||||| ||||||||||| ||||| ||| ||||||||||||
Db   229 CCAAGATATGTTGTTGCGTTGTGTCGTCGACAGCTTGACAAAAGTAAGCCACCAGAACA 288

Qy   370 AGAGTTGCACTCCGCAACGAACAGTTATGTTCTTCACTTAGTTCGTGCCAAGTAAAGTTGAT 429
           |||||||| |||||| || |||| ||| ||||||| |||| ||||| ||| |||||||
Db   289 AGAGTTGCTTTGGATATGACTACACTAACTATCATCAGAGATATTTGCCGAGCAGAGGGGAT 348

Qy   430 CCATTGGTCAAACTGATGAAAGTTGACAAAGTTCCCGATTCTTACACATATGACTGACTGGT 489
           |||| ||| ||| ||||||| | ||| || ||||||| || |||||| |||||||||
Db   349 CCACTGGTTTATAACATGTCTCATGACCCCTGGGAAATGTTTCTTATTCTCGAGATTGGA 408

Qy   490 GGTTTTAGACCAGCAAATTAAAGAAAATAAAAGAGGTCATTGTAGCTACAAAATCAAACATTCCT 549
           ||| |||| || ||||||||| |||||||| |||||||||||| || |||||||||||
Db   409 GGGCTACGACGATCCGGGAATTAAGAGAGGTGATGAGAATTACCTCTTACAAACCCA 468

Qy   550 GAGCTGTTTGAAAGTCTTGGAATGCACAACCAAAGGGTGTCCTGCTCTAGGGCCACCT 609
           ||||||||| |||| |||||||||| || ||||||||| || |||||||| || |||
Db   469 GAGTATTTCACGCTGTGAGGAATAATACCTCCAAAAGCGTGTTGTTTATATGGACCACCA 528

Qy   610 GGTACAGGTAAAACATTCTTGGGTTAGGGCCAGTGGCCTCATCATACTGACTGTACATTCATC 669
           |||| ||||||| ||| || |||| ||| ||| || || ||||||||||||||||
Db   529 GGTACGGGGAAAAACACTCTTGGGCACGAGACCGTTGCTAGCCAGCTGGACTGCAATTCTTA 588

Qy   670 AGGGTGTCTCGGTTCTGAGTTAGTTCAGAAAATACATTGGAGAAGAGGTTCTTAGAATGGGTCAGG 729
           || | |||| || || || || |||| || |||||||||| || |||||||||||||
Db   589 AAGGTGTATCTAGTTCTATTGTAGACAAGTAGACTGTGAAAAGTGCTCGTTTGATCAGA 648

Qy   730 GAACTTTTTGTTTATGGCCAGGGAACATGCTTCCATCAATTAATCTTCATGGATGAAAATTGAC 789
           |||| ||| |||||||||| || |||| |||||||||| ||||| ||||||||
Db   649 GAAAIGTTTAATTATGTGTAGAAGATCATCAACCATGCATCATTTTATAGGATGAAAATAGAT 708

Qy   790 AGTATGGGATCTCGTCCTCCGGATGGAATCTGGAAGTGGCAACGGTGATGAGTGAGGTACGGCGT 849
           ||||| |||| || |||||||||| || |||||| ||||||| |||||||||| ||
Db   709 GCTATTGGGATCGGTCGTCGGGTTT---TCTGAGGGTACTTCAGCTGACACAGAGAGAATTCAGAGA 765

Qy   850 ACTATGCTGGAACTTCTCAACCAGTGGATGGATTTGAAGCTTCAAATAAGATCAAGGTT 909
           |||| || |||||||| || |||| |||||| ||||||| ||||||| |||||||
Db   766 ACGTTAATGGAGTTACTGAATCAAAATGGATAGGGATTTGATACTCGCAGAGAGTTAAAATG 825

Qy   910 TTGATGGCCACCAATCGGATTGATATCCTGGATCAAGGCCTCTTAGACCAGAGACGGATA 969
           ||||||||||||||||| ||||| ||||||||||||||| |||||||||||||| |||
Db   910 TTGATGGCCACCAATCGGATTGATATCCTGGATCAAGGCCTCTTAGACCAGAGACGGATA 969

RESULT 8
US-08-820-170A-15
; Sequence 15, Application US/0882017OA
; Patent No. 5831058
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIWARA
; APPLICANT: Takeshi, WATANABE
; APPLICANT: Masato, HORIE
; TITLE OF INVENTION: Toyomasa, KATAGIRI
; NUMBER OF SEQUENCES: 42
; TITLE OF INVENTION: HUMAN GENE
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/820,170A
; FILING DATE:
; CLASSIFICATION:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 536
; TELEFAX: (202) 293-7060
; TELEX: 6491103
; TELEPHONE: (202) 293-7860
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1566 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; LIBRARY: Human fetal brain cDNA library
; CLONE: GEN-331G07
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 17..1183

US-08-820-170A-15

Query Match   21.6%;   Score 270.6; DB 2; Length 1566;
Best Local Similarity   55.9%;   Pred. No. 2.9e-74;
Matches   535; Conservative   0; Mismatches   419; Indels   3; Gaps   1;

Qy   250 GTCGGTGAAGTGTCAAAGTAATGGGCAAGAACAAAGTCCTTGTCCACTGACTCCACCAGAA 309

; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/055,699
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/820,170
; FILING DATE:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 649103
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1566 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; LIBRARY: Human fetal brain cDNA library
; CLONE: GEN-331G07
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 17..1183
; US-09-055-699-15

Query Match   21.6%;   Score 270.6; DB 3;   Length 1566;
Best Local Similarity   55.9%;   Pred. No. 2.9e-74;
Matches 535; Conservative   0; Mismatches 419; Indels   3; Gaps   1;

US-09-055-699-15

RESULT 9
US-09-055-699-15
; Sequence 15, Application US/09055699
; Patent No. 6005088
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIWARA
; APPLICANT: Takeshi, WATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI

US-09-273-565-15

```
Qy   670  AGGGTGTCGTTCTGAGTTAGTTCAGAAATACAHTGGAGAATGGTCAGG  729
          |||||||
Db   605  AAGGTTGTATCTAGTTCTATTGTAGACAAGTACATTGGGAAAGTGCTTCGGTTTGATCAGA  664

Qy   730  GAACTTTTTGTTATTGGCACGGGGAACACATGCTCCATCAATTATTACTTCAGGATGAAATTGAC  789
          ||||
Db   665  GAAAATGTTTAATTATGCTAGAGATCATCAACCAATGCCATCAATTTTATGGATGAAATAGAT  724

Qy   790  AGTATTGGAATCTGCTCGGATTGGAAATCTGGAAGTGGGCAACGCGGTAGTTAGTGAGGTACAGGCT  849
          ||||||
Db   725  GCTATTGGTGGTCGTCGGTTT---TCTGAGGGTACTTCAGCTGACAGAGAGATTCAGAGA  781

Qy   850  ACTATGCTCGGAACTTCTCAACCAGTTGGATTGGAGCTTCAAATAAGACAAGGTT  909
          ||||
Db   782  ACGTTAATGGAGTTACTGAGATCAAATGGATTGGATAGATACTCATAGAGAGTTAAAATG  841

Qy   910  TTGATGTGGCCACCAATCGGATTGATATCCTGGATCAAGCCCTCCTTAGACCAGACCGGATA  969
          ||||
Db   842  ACCATGGCTACAAACAGACCAGATCCTGCTTTGCCGTCCGTTTCATCAGAGAGTTAAAAATC  901

Qy   970  GACCGGAAAATTGAATTCCAACCCCTAATGAAGAAGTCTCGGCGCATTGAAGAAGATTGCCCGAGAAGATG  1029
          ||||
Db   902  GATAGAAAAATACATATTGATTGGCCAAATGACACACTGAACATTAGACATTTTGAAAATC  961

Qy  1030  CATTCTAGAGAAGAATGAAATTAATGCGGTGGCCATTGAATTGAAAGAAGAATTGCCGAGGAAGATG  1089
          ||||
Db   962  CATGGCAGGTCCCATTACAAAGCATGGTGAAAATAGATTAATAGAGAAGCAAATTGTGTTCTTCG  1021

Qy  1090  AATGGAGAGCAATCTGGTGGTCCTGCGAACTTAAGGCCTGTTCGCACTGAAGATGTTTTGCTTTG  1149
          ||||
Db  1022  GATGGCTTTAATGGAGCAGATCTCAGAAAATGTTGTACTGAGGCAGGTATGTTCGCAATT  1081

Qy  1150  AGGGAGCGGAGGGTACACGTGACTCAGCAGGATTTTGAGATGGCCGTGGCGAAGGTG  1206
          ||||
Db  1082  CGTGCTGATCATGATTTTGTAGTTACAGGAAGACTTCATGAAAGCAGTCAGAAAAGTG  1138
```

RESULT 10
Sequence 15, Application US/09273565A
; Patent No. 6166190
; GENERAL INFORMATION:
; APPLICANT: FUJIWARA, TSUTOMU
; APPLICANT: WATANABE, TAKESHI
; APPLICANT: HORIE, MASATO
; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
; TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
; FILE REFERENCE: Q-53599
; CURRENT APPLICATION NUMBER: US/09/273,565A
; CURRENT FILING DATE: 1999-03-22
; EARLIER APPLICATION NUMBER: 09/055,699
; EARLIER FILING DATE: 1998-04-07
; EARLIER APPLICATION NUMBER: 08/820,170
; EARLIER FILING DATE: 1997-03-19
; EARLIER APPLICATION NUMBER: JP 63410/1996
; EARLIER FILING DATE: 1996-03-19
; EARLIER APPLICATION NUMBER: JP 69163/1997
; EARLIER FILING DATE: 1997-03-05
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 15
; LENGTH: 1566
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (17)..(1183)
US-09-273-565-15

Query Match   21.6%;
Best Local Similarity   55.9%;
Matches 535; Conservative 0; Mismatches 419; Indels 3; Gaps 1;

Score 270.6; DB 3; Length 1566;
Pred. No. 2.9e-74;

```
Qy   250  GTCCGGTGAAGTTGTCAAAGTAATGGGCAACAGAAACAAAGTCCTTGGTCACCCAGAA  309
          |||||
Db   185  GTGGGTGAAGTGCTTCAGTTGCTTAAACAGTTGAACATGAAGAAAAATTCATTGGAATGGA  244

Qy   310  GGAAAAATATGTTGACATTGACAAAAATATTGACAATTACACTTCCATCCACT  369
          |||||
Db   245  CCAAGATAATGTTGTGGGGTTTGTCGTCGCACGCTTGACAAAAGTAAGCCAGGACA  304

Qy   370  AGAGTTGCACTCCGCAAGGACACAGTTATGTTCTTCACTTAGTTCTGCCAAGTAAAGTTGAT  429
          |||||
Db   305  AGAGTTGCTTTGGATATGACTAACAACTATCATGAGAGTATATTTGCCGAGAGAGGTGGAT  364

Qy   430  CCATTGGTCAAATCTGATGAAAGTTGAGAAAGTTCCGGATTCAAATAAGACAAGGTT  489
          |||||
Db   365  CCACTGGTTTATAACATGTCATCAGAGAGAACCCTGGAAGTGTTTCTTAGGATGGA  424

Qy   490  GGTTAGACCAGCAAAATTAAAGCAAATAAAAGAGGTCAATTCATTGAGCTACCAAGTCAAACATCCT  549
          |||||
Db   425  GGGCTATCAGAACAGATCCGGGAATTAAGAGAGTTCCAGTCGATGAAGTTCCGGA  484

Qy   550  GGTACGGGGAAAAACACTCTTGGCACGAGGTGGCCTCATCATACTGACTGTACATTCATC  609
          |||||
Db   485  GAGTTATTTCAGCGGTGTAGGAATAAATACCTCCAAAAAGGGTTCGCCACCT  544

Qy   610  GGTACAGGTAAAAACATTGTTGGCGTAGGCGCAGTGGCCTCATCATACTGACTGTACATTCATC  669
          |||||
Db   545  GGTACGGGGAAAAACACTCTTGGCACGAGGCCGCTTGCTAGCCAGTGCAATTTCTTA  604

Qy   670  AGGGTGTCTGGTTCTGAGTTAGTTCAGAAATACAHTGGATGAATGGTCAGG  729
          |||||
Db   605  AAGGTTGTATCTAGTTCTATTGTAGACAAGTACATTGGGAAAGTGCTTCGGTTTGATCAGA  664

Qy   730  GAACTTTTGTTATTGTTATGGCCAGGAACTTAAGCCTGTCTGCCAGGAGAAGATTGAC  789
          |||||
Db   665  GAAAATGTTTAATTATGCTAGAGATCATCAACCAATGCCATCAATTTTATGGATGAAATAGAT  724

Qy   790  AGTATTGGAATCTGCTCGGATGATCGAAGGTGGCCAACGCGGTAGTTAGTGAGGTACAGGCGT  849
          |||||
Db   725  GCTATTGGTGGTCGTCGGTTT---TCTGAGGGTACTTCAGCTGACAGAGAGATTCAGAGA  781

Qy   850  ACTATGCTCGGAACTTCTCAACCAGTTGATGGATTCTCAAATAAGACAAGGTT  909
          |||||
Db   782  ACGTTAATGGAGTTACTGAGATCAAATGGATTGGATAGATACTCATAGAGAGTTAAAAATG  841

Qy   910  TTGATGGCCACCAATCGGATTGATCGAATATCCTGGATCAAGCCCTCCTTAGACCAGACGACGGATA  969
          |||||
Db   842  ACCATGGCTACAAACAGACCAGATCCTGCTTTGCCGTCCGTTTCATCAGAGAGTTAAAAATC  901

Qy   970  GACCGGAAAATTGAATTCCAACCCCTAATGATGAAGTCTCGGCGTCTGGATATTTTTGAAAATC  1029
          |||||
Db   902  GATAGAAAAATACATATTGATTGGCCAAATGACAACAGTGAAACATACTACACTGAAAATGGA  961

Qy  1030  CATTCTAGAGAAGAATGAAATTAATGCGTGGCATTGATTTGAAAGAAGAAGATTGCCGAGGAAGATG  1089
          |||||
Db   962  CATGCCAGGTCCCATTACAAAGCATGGTGAAATAGATTATGAAGCAAATTGTGAAGCTTTCG  1021

Qy  1090  AATGGAGCATCTCGGTGCTGGAACTTAAGGCCTGTTGCACTGAAGGTGGAATTGTTGCTTTG  1149
          |||||
Db  1022  GATGGCTTTAATGGAGCAGATCTCAGAAAATGTTGTTACTGAGGCAGGTATGTTCGCAATT  1081

Qy  1150  AGGGAGCGGAGGGTACACGTGACTCAGCAGGATTTTGAGATGGCCGTGGCGAAGGTG  1206
          |||||
Db  1082  CGTGCTGATCATGATTTTGTAGTTACAGGAAGACTTCATGAAAGCAGTCAGAAAAGTG  1138
```

RESULT 11
US-09-565-538-15
; Sequence 15, Application US/09565538
; Patent No. 6333404
; GENERAL INFORMATION:
; APPLICANT: FUJIWARA, TSUTOMU
; APPLICANT: WATANABE, TAKESHI

; APPLICANT: HORIE, MASATO
; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
; TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
; FILE REFERENCE: Q-53599
; CURRENT APPLICATION NUMBER: US/09/565,538
; CURRENT FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: 09/273,565
; PRIOR FILING DATE: 1999-03-22
; PRIOR APPLICATION NUMBER: 09/055,699
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: 08/820,170
; PRIOR FILING DATE: 1997-03-19
; PRIOR APPLICATION NUMBER: JP 63410/1996
; PRIOR FILING DATE: 1996-03-19
; PRIOR APPLICATION NUMBER: JP 69163/1997
; PRIOR FILING DATE: 1997-03-05
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 15
; LENGTH: 1566
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (17)..(1183)
US-09-565-538-15

Query Match   21.6%;   Score 270.6; DB 4; Length 1566;
Best Local Similarity 55.9%;   Pred. No. 2.9e-74;
Matches 535; Conservative 0; Mismatches 419; Indels 3; Gaps 1;

```
Qy  250  GTCGGTGAGTTGTCAAAGTAATGGCAAGAACAAAGTCCTTGTCAAGGTCCACCCAGAA  309
                                                                       
Db  185  GTGGGTGAAGTGCTTAAACAGTTAACTGAAGAAAAAATTCATTGTTAAAGCTACCAAATGGA  244

Qy  310  GGAAAATATGTTGTTGACATTGACAAAAATATTGACATTACAAAGATTACTCCATCCACT  369
                                                                       
Db  245  CCAAGATATGTTGTGGGTTCGTCGTGACAGCCTTGACAAAGTAAGCTGAAGCCAGGAACA  304

Qy  370  AGAGTTGCACTCCGCAACGACAGTTAGTGTTCTTCACTTAGTTCTGCCAAGTAAAGTTGAT  429
                                                                       
Db  305  AGAGTTGCTTTGGATATAGTACTACTAATCATGAGATATTTGCCGAGAGAGGTGGAT  364

Qy  430  CCATTGGTCAAATCTGATGAAAAGTTGAGAAAGTTCCCGATTCTACATATGACATGATTGGT  489
                                                                       
Db  365  CCACTGGTTTATAACATGTCTCATGAGGACCCTGGGAATGTTTCTTATTCTGAGATTGGA  424

Qy  490  GGTTTAGACCAGCAAAATTAAAGAAATAAAAAGAGGTCATTGAGCTACCAAACATCCT  549
                                                                       
Db  425  GGGCTATCAGAACAGAATCCCGGGAATTAAGAGGTGAATAGAATTACCTCCTTACAAACCCA  484

Qy  550  GAGCTGTTTGAAAGTCCTTGGAATTCACAACCAAAGGGGTGTCCTGCTCTATAGTGGCCACCT  609
                                                                       
Db  485  GAGTTATTTCAGCCGTGTAGGAAATAATACCTCCAAAAGGCTGTTTGTTTATGACCACCA  544

Qy  610  GGTACACAGGGTAAAACAATTGTTGGGCTGCTAAGGCAGTGGCTCATCAGTGACTGACTACAATCATC  669
                                                                       
Db  545  GGTACGGGGAAAAACACTCTTGGCACGAGCCCGTTGCTAGCCAGCTGGACTGCAATTCTTA  604

Qy  670  AGGGGTGTCTGGGTTCTGAGTTAGTTCAGGAAATACATTGGGAGAAGGTTCTTAGAATGGTCAGG  729
                                                                       
Db  605  AAGGTGTATCTAGTTCTATTGTAGGACAAGTACATCAATTGGTTGGAAAAGTGCTCGGCTTTG  664

Qy  730  GAACTTTTGTTATGGCACGGGACAGCATGCTCCATCAAATATCTTCATGGATGAAATTGAC  789
                                                                       
Db  665  GAAAGTTTAATTATGCTAGAGATCATCAACCATGCATATTTTTTATGGATGAAATAGAT  724

Qy  790  AGTATTGGATCTCGTCGCTGGATTGGAATCTCGAAGTGGCAACGGGTACATGGGTAGTGCAGGGTACAGCGCT  849
                                                                       
Db  725  GCTATTGGGTGGTTCGGTCGGTCGGTTT---TCTGAGGGTACTTCAGCTGACAGAGGATTCAGA  781

Qy  850  ACTATGCTCGGAACTTCTCAACCCAGTTGGATTGGAATGGATTTGAAGCTTCAAAATAGAGATCAAGGTT  909
```

RESULT 12
US-09-661-468-15
; Sequence 15, Application US/09661468
; Patent No. 6376189
; GENERAL INFORMATION:
; APPLICANT: FUJIWARA, TSUTOMU
; APPLICANT: WATANABE, TAKESHI
; APPLICANT: HORIE, MASATO
; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
; TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
; FILE REFERENCE: Q-53599
; CURRENT APPLICATION NUMBER: US/09/661,468
; CURRENT FILING DATE: 2000-09-13
; PRIOR APPLICATION NUMBER: 09/055,699
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: 08/820,170
; PRIOR FILING DATE: 1997-03-19
; PRIOR APPLICATION NUMBER: JP 63410/1996
; PRIOR FILING DATE: 1996-03-19
; PRIOR APPLICATION NUMBER: JP 69163/1997
; PRIOR FILING DATE: 1997-03-05
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 15
; LENGTH: 1566
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (17)..(1183)
US-09-661-468-15

Query Match   21.6%;   Score 270.6; DB 4; Length 1566;
Best Local Similarity 55.9%;   Pred. No. 2.9e-74;
Matches 535; Conservative 0; Mismatches 419; Indels 3; Gaps 1;

```
Qy  250  GTCGGTGAGTTGTCAAAGTAATGGCAAGAACAAAGTCCTTGTCAAGGTCCACCCAGAA  309
                                                                       
Db  185  GTGGGTGAAGTGCTTAAACAGTTAACTGAAGAAAAAATTCATTGTTAAAGCTACCAAATGGA  244

Qy  310  GGAAAATATGTTGTTGACATTGACAAAAATATTGACATTACAAAGATTACTCCATCCACT  369
                                                                       
Db  245  CCAAGATATGTTGTGGGTTCGTCGTGACAGCCTTGACAAAGTAAGCTGAAGCCAGGAACA  304

Qy  370  AGAGTTGCACTCCGCAACGACAGTTAGTGTTCTTCACTTAGTTCTGCCAAGTAAAGTTGAT  429
                                                                       
Db  305  AGAGTTGCTTTGGATATAGTACTACTAATCATGAGATATTTGCCGAGAGAGGTGGAT  364

Qy  430  CCATTGGTCAAATCTGATGAAAAGTTGAGAAAGTTCCCGATTCTACATATGACATGATTGGT  489
```

RESULT 13
US-09-976-165-15
; Sequence 15, Application US/0976165
; Patent No. 6562947
; GENERAL INFORMATION:
; APPLICANT: FUJIWARA, TSUTOMU
; APPLICANT: WATANABE, TAKESHI
; APPLICANT: HORIE, MASATO
; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
; TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
; FILE REFERENCE: Q-53599
; CURRENT APPLICATION NUMBER: US/09/976,165
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 09/565,538
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: 2000-05-05
; PRIOR APPLICATION NUMBER: 09/055,699
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: 08/820,170
; PRIOR FILING DATE: 1997-03-19
; PRIOR APPLICATION NUMBER: JP 63410/1996
; PRIOR FILING DATE: 1996-03-19
; PRIOR APPLICATION NUMBER: JP 69163/1997

; PRIOR FILING DATE: 1997-03-05
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 1566
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (17)..(1183)
US-09-976-165-15

Query Match   21.6%;   Score 270.6; DB 4; Length 1566;
Best Local Similarity   55.9%;   Pred. No. 2.9e-74;
Matches 535; Conservative   0; Mismatches 419; Indels   3; Gaps   1;

RESULT 14
US-08-998-416-600/c
; Sequence 600, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 62392264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/1/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 600:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 727 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGINAL SOURCE: PAG1410UP
US-08-998-416-600

Query Match   16.2%;   Score 203.2; DB 3;   Length 727;
Best Local Similarity   59.8%;   Pred. No. 2.1e-53;
Matches   359;   Conservative   0;   Mismatches   238;   Indels   3;   Gaps   1;

Qy   412   CTGCCAAGTAAAGTTGATCCATTGGTCAATCTGATGAAAAGTTGAGAAGTTCCCGATTCT 471
Db   606   CTTCCTTCCGAGTTTGACTCTAGAGTGAAAGCTATGGAAGTGAATGGATGACAAGCTACGGAA 547
Qy   472   ACATATGACATGATTGGGGTTGGTTTAGACCAGCAAATTAAAGAAAATAAAAGAGGTCATTGAG 531
Db   546   ACATACTCCGACCGTTGGGTGGGCTAGACAAGCAGATCGAGGAACTAGTCGGAAGCAAATCGTG 487
Qy   532   CTACCAATCAAAACATCCTAGCTCGTTGAAAGTCTTGGAATTGCACAACCAAAGGGGTGTC 591

RESULT 15
US-09-313-294A-1050
; Sequence 1050, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS.: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 1050
; LENGTH: 277
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6476212 7005503051H1
US-09-313-294A-1050

Query Match   15.3%;   Score 192.2; DB 4;   Length 277;
Best Local Similarity   80.9%;   Pred. No. 3.2e-50;
Matches   224;   Conservative   0;   Mismatches   53;   Indels   0;   Gaps   0;

Qy   845   AGGGTACTATGCTTGGGACTTCTCCAACCAGTTGGATGGATTTGAAGCTTCAAATAAGATCA 904
Db   1   AGCCTACTATGCTTGAACTTCCTAAACCAGCTCGATGGTTTTGAAGGCATCAAACAACAAAATTA 60
Qy   905   AGGTTTTGATGGCCACCAATCGGATTGATATATCCTGGAATCAAAGCCCTCCTTAGACCAGAC 964
Db   61   AGGTTTTGAATGGCAACGAACAGAATAGACAATTTTGGATCAAGCGCCTTCTGAGCGCCTGGCC 120
Qy   965   GGATAGACCGGAAATTGAATTCCACCCCTAATGAAGAGTGTCTCGGCTGGATATTTGA 1024
Db   121   GCATAGACAGGAAGATTGAAATTCCAAAATCCTAACGAGAGGATTCACGTTTCGATATCTGA 180
Qy   1025   AAATCCATTCTAGAAGAATGAATTAATGCGTGGCATTGAATTTGAAGAAGAGTTGCCGAGA 1084

```
Db   181 AGATCCCATTCAAGAAAAATGAACTTGATGCGTGGCATTGATCTGAAAAAGATCGCGGAAA 240
Qy  1085 AGATGAATGGAGCAATCTCGGTGCTGAAACTTAAGGCCTGT 1121
          |||||||||| || || ||  ||||||| ||
Db   241 AGATGAATGGGGCCTCAGGAGCTGAGCTTCAAGGCCCGT 277
```

Search completed: November 9, 2003, 11:16:59
Job time : 116 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: November 9, 2003, 10:24:31 ; Search time 475 Seconds
                8424.128 Million cell updates/sec
                (without alignments)

Title: US-09-462-972-1
Perfect score: 1254
Sequence: 1 atggctctgtaggagttcga..........cattgcggaagtcgtggaag 1254

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched: 2141354 seqs, 1595478879 residues

Total number of hits satisfying chosen parameters: 4282708

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Database : Published Applications NA:*
  1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
  2: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
  3: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
  4: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
  5: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
  6: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
  7: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
  8: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
  9: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
 10: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
 11: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
 12: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
 13: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
 14: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
 15: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
 16: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
 17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match length	DB	ID	Description
1	633.4	50.5	1505	9	US-09-925-301-142	Sequence 142, Appl
2	536.6	42.8	1206	12	US-10-032-585-6325	Sequence 6325, Ap
3	510.4	40.7	1170	14	US-10-128-714-2369	Sequence 2369, Ap
4	510.4	40.7	1170	14	US-10-128-714-7369	Sequence 7369, Ap
5	324.2	25.9	1351	14	US-10-128-714-1369	Sequence 1369, Ap
6	324.2	25.9	1351	14	US-10-128-714-6369	Sequence 6369, Ap
7	324.2	25.9	3350	14	US-10-128-714-369	Sequence 369, Appl
8	324.2	25.9	1351	14	US-10-128-714-5369	Sequence 5369, Ap
9	304.6	24.3	1332	10	US-09-938-842A-1753	Sequence 1753, Ap
10	285.8	22.8	1335	12	US-10-032-585-6385	Sequence 6385, Ap
11	278	22.2	1566	12	US-10-205-219-166	Sequence 166, App
12	275.2	21.9	1398	12	US-10-032-585-6680	Sequence 6680, App
13	272.2	21.7	1891	12	US-10-101-510-379	Sequence 379, App
14	271.8	21.7	1320	14	US-10-128-714-2522	Sequence 2522, Ap
15	271.8	21.7	1320	14	US-10-128-714-7522	Sequence 7522, Ap
16	270.6	21.6	1167	10	US-09-976-165-14	Sequence 14, Appl
17	270.6	21.6	1167	12	US-10-342-276-14	Sequence 14, Appl
18	270.6	21.6	1566	10	US-09-976-165-15	Sequence 15, Appl
19	270.6	21.6	1566	12	US-10-342-276-15	Sequence 15, Appl
20	263.6	21.0	1182	14	US-10-128-714-2421	Sequence 2421, Ap
21	262	20.9	1356	14	US-10-128-714-7421	Sequence 7421, Ap
22	258.2	20.6	1287	12	US-10-032-585-6464	Sequence 6464, Ap
23	255.6	20.4	1490	12	US-10-128-714-1522	Sequence 1522, Ap
24	255.6	20.4	1490	14	US-10-128-714-6522	Sequence 6522, Ap
25	255.6	20.4	3490	14	US-10-128-714-522	Sequence 522, App
26	255.6	20.4	3490	14	US-10-128-714-5522	Sequence 5522, Ap
27	250.8	20.0	1392	14	US-09-878-574-456-509	Sequence 509, App
28	249.2	19.9	1341	10	US-09-878-574-3203	Sequence 3203, A
29	249.2	19.9	1341	14	US-10-171-581-6	Sequence 6, Appli
30	249	19.9	269	14	US-09-878-574-20461	Sequence 20461, A
31	248	19.8	502	11	US-09-917-800A-1340	Sequence 1340, A
32	246.6	19.7	1397	10	US-09-918-995-20461	Sequence 20461, A
33	243.4	19.4	274	14	US-10-128-714-2418	Sequence 2418, Ap
34	242.8	19.4	1287	14	US-10-128-714-5757	Sequence 5757, Ap
35	242.8	19.4	1395	14	US-10-128-714-7418	Sequence 7418, Ap
36	241.4	19.3	505	14	US-10-128-714-961-328	Sequence 328, App
37	238.4	19.0	260	10	US-09-770-961-12353	Sequence 12353, A
38	237	19.0	1293	10	US-09-878-574-12353	Sequence 12353, A
39	235	18.7	1398	14	US-10-032-585-6453	Sequence 6453, Ap
40	232	18.5	1266	14	US-10-128-714-2485	Sequence 2485, Ap
41	227.6	18.1	477	10	US-09-918-995-5993	Sequence 5993, Ap
42	221.4	17.7	274	12	US-10-128-714-7485	Sequence 7485, Ap
43	220.8	17.6	1236	12	US-10-032-585-6284	Sequence 6284, Ap
44	206.6	16.5	1235	14	US-10-128-714-1256	Sequence 1256, Ap
45	206.6	16.5	1596	14	US-10-128-714-6256	Sequence 6256, Ap

ALIGNMENTS

RESULT 1
US-09-925-301-142

Sequence 142, Application US/09925301

Patent No. US2002005230A1

GENERAL INFORMATION:
  APPLICANT: Rosen et al.
  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
  FILE REFERENCE: PA106
  CURRENT APPLICATION NUMBER: US/09/925,301
  CURRENT FILING DATE: 2001-08-10
  PRIOR APPLICATION NUMBER: PCT/US00/05882
  PRIOR FILING DATE: 2000-03-08
  PRIOR APPLICATION NUMBER: 60/124,270
  PRIOR FILING DATE: 1999-03-12
  NUMBER OF SEQ ID NOS: 1694
  SOFTWARE: PatentIn Ver. 2.0
  SEQ ID NO 142
  LENGTH: 1505
  TYPE: DNA
  ORGANISM: Homo sapiens
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (1193)
  OTHER INFORMATION: n equals a,t,g, or c
  NAME/KEY: misc_feature
  LOCATION: (1499)
  OTHER INFORMATION: n equals a,t,g, or c
  NAME/KEY: misc_feature
  LOCATION: (1500)
  OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-142

Query Match   50.5%; Score 633.4; DB 9; Length 1505;
Best Local Similarity   71.7%;
Matches 845; Conservative 0; Mismatches 331; Indels 3; Gaps 1;

Qy   76 AAGCAGGGCGAGGGCCTCCGCCACTACTATTCTCTCAACATCCACGAGCATCAGCTCCTT 135
         ||| ||| |||| |||| || ||| ||||| | || | || || ||| | | || |||
Db 208 AAGGCAGGCGAGCGGACTCCGCCAATATTATCTGTCCAAGATTGAAGAACTCCAGCTGATT 267

RESULT 2
US-10-032-585-6325

; Sequence 6325, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6325
; LENGTH: 1206
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-6325

Query Match   42.8%; Score 536.6; DB 12; Length 1206;
Best Local Similarity   66.6%; Pred. No. 2.8e-150;
Matches 783; Conservative   0; Mismatches 389; Indels   3; Gaps   1;

Query Match   40.7%;   Score 510.4;   DB 14;   Length 1170;

RESULT 3

US-10-128-714-2369

; Sequence 2369, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRICR APPLICATION NUMBER: US 60/316,362
; PRICR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2369
; LENGTH: 1170
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-2369

Best Local Similarity   65.4%;   Pred. No. 2.1e-142;
Matches 764;   Conservative   0;   Mismatches 401;   Indels   3;   Gaps   1;

Qy   739 GTTATGGCCAGGAGGAGCACTGCTCCATCAATTATCTTCATGGATGAAATTGACAGTAGTTGGA 798
Db   691 GTGATGGCAAGGAGGACAATGGCACCAATCAATCATTTTATGGATGAAATTGAATTCTATTGG 750

Qy   799 TCTCCTCGGATGGAATCTCGGAAGTTGGCAACGGTGGATAGTAGTTACACGGTACTATGCTG   858
Db   751 TCGTCGGCGTCGTTGNAGGTTCATCAGG---AGGTGATTCCGAAGTGCAAAGAACAAATGTTG   807

Qy   859 GAACTTCTCAACCAGTTGGATGATGGAATTGAAGCTTCAAAATAAGATCAAGCGTTTTGATGGCC   918
Db   808 GAATTGTTAAATCAATTGGATGGGTTTGAAAGTTCAAAGATAATTCAAAATCATTCATATGGCC   867

Qy   919 ACCAATCGGATTGAATATCCTGGATCAAGGCCTCCTTAGGACCAGGAACGGATAGAACCGGAAA   978
Db   868 ACTAATCGTTTAGATATTTGGATCCAGCTTTATTGAGACCAGGTAGAAATTGAATCAAGA   927

Qy   979 ATTGAATTTCCAACCCCTAATGAAGAGTCTCGCGTGGCATTGAAGAGTCCTTGTCAAGGTTCTAGA 1038
Db   928 ATTGAATTCCCAGCACCAACCGTTGCGCCTAGAACAGATATTTAAGATTCATTCAAGA   987

Qy 1039 AGAATGAATTTAATGCGTGGCATTGATTTGAAGAGATTGCCGAGAAGGATGAATGGAGCA 1098
Db   988 TCCATGAATTTGACAAGGAGTTATTAATTGCGTAAAAATTGGAGAAAATGAATGGAATGC 1047

Qy 1099 TCTGGTGGCTGAACTTAAGGCTGTTTGCACTGAAGCTGGAAATGTTTGCTTGAGGGAGGCGGG 1158
Db 1048 AGTGGTGCTGATGGAAAAGGAGTTTGGTACCGGAAGCTCGGTATGTTATGCATTGAGCCTTAT 1107

Qy 1159 AGGGTACACGTGACTCAGGAGGATTTGAGATGGCCGTGGGCCGAAGGTGGATTCGGCAGTGAAAAAGGAG 1218
Db 1108 AGAATTCATTGTAACTCAAGAGAGATTTCGAATTGGCAGTTGCAAAAGTCAATGTCTAAGGAT 1167

Qy 1219 ACTGAAAAAAACATGTCATTGCCGAAGTTGTGGAA 1253
Db 1168 GACGATGGTGCGTGTCTCCTTACAGAAAATTATTCAA 1202

Qy   87 GGGCCTTCCGCCACTACTTACTTCTTCATTCTCTCACATCCACAGGCGTCTGGGTTCTGA 146
Db   3 GGGCCTGGACAACTTACTTATCGCAATAAGATCCGAGAGTATGAAAACTCAGCTCCTTCTTTGCCAAAA   62

Qy   147 GACTNCATAACCTCAAACCGTCTCGAGGCCTCGAGAGAAAACGACCTCA.TTCTAGGGTGAGGAT 206
Db   63 CCAGGCCTGTTTTACGTCGACTGAAGCACAGCGCAACGACTAATTCGGCGAGTTCGCCT 122

Qy   207 GCTGCGCGGAAGAATTACAGCTTCTGCAGGAACCCGGGTCCTTTATGTCGGTGAAGTGTCAA 266
Db   123 ACTACGAGGAGGAGTTGGGCTTGCTTCACAACCTGGGTCTTATGTGGGAGAGGTCGTTGA 182

Qy   267 AGTAATGGGCGAGAACAAAGTCCTTGTCAAGGTCCAGCCCACCCGGAGGAAAATATGTTGTTGA 326
Db   183 GGTGATGAGCACAAGAAGGTTCTTGTCAAGGTGACTACCGAGAACGAGAAATATGTCGTTGA 242

Qy   327 CATTGACAAAAATATTTGAACATTACAAAGATTACTACTCCATCCACTAGAGTTGCACTCCGCAA 386
Db   243 CAATTGCAGACGCGGTCGATATTACGAAAACTTACTGTGGGCAAGCGTGTTGCTTTGCTTCC 302

Qy   387 CGACAGTTATGTTCTTCACTTAGTTCTGCCCAAGTAAAGTTGATCGTGGTCAAATCTGAT 446
Db   303 AGACTCTTATAATTAGAAAAATGTTGCCATCGTCGGTTTGACCGGTCGTTTCGCCTTAT 362

Qy   447 GAAAGTTGAGAAAGTTCCCGATTCTACAATATGACATGATTGGTTGGTTGATTGAGACCAGCAAAT 506
Db   363 GATGGTCGAGAAGAGGTGCCTGACCATTACGGATAGGCGATATGGCCTTGACCAACAAAT 422

Qy   507 CAGGGAACATGCTCCATCAATTATCTTCATGGATGAAAATTGACAGTAGTTGGATCTGCTG 566
Db   423 CAAGGAGATCAAGGAAGTCATTGAGCTTGGTGTCTCAACCATCCAAGAACTGTTCGGAGTCTCT 482

Qy   567 TGGAATTGCACACACCAAAAGGGTGTTCTCTTTTACGGCGGGCCCACTCGGTACAGGTAAAAACAT 626
Db   483 TGGTATCGCACAGCCCAAAAGGTGTTCTTTCTTTTACGGGGAACCGGGTAAAACACT 542

Qy   627 GTTGGGCTAGGGCCAGTGGCCTCATCATTACTACTGTACATTCACAGGGTGTCTGGTTCTGA 686
Db   543 GCTCGCCCGAGCCGGTAGCCCATCACACAGATGCCAAGGTGCATCGCTCGGA 602

Qy   687 GTTAGTTCAGAAATACATTGGAGAAGGTTCTAGAATGGTCAGGGAACTTTTTGTTATGGC 746
Db   603 ACTAGTCCAGAAATACATTGGTGAGGCTAGTCGCATGGGCTAGTTTCGTTCGTCATGGC 662

Qy   747 CAACCAGTTGATGGAATTGAAGCTTCAAAATAAGATCAAGGTTTTGATGGCCACCAATCG 806
Db   663 CAATCAGCTGGATGGGGTTGAGCCCAAAAAACATCAAAATCATGGCTAGCAACCG 722

Qy   807 GATGGAATCTGGAAGTGGCAACGGTGGATAGTGAGGTGGTACAGGGTACTACTTGTCTGAACTTCT 866
Db   723 TATAGACTCGG---CTGGCTCTGGAGAGATTCAGGAGTGCGGTCGTCATATGTGGGACTGCT 779

Qy   867 CAACCAGTTGGATGATGGAATTGAAGCTTCTAGAAATAAGATCAAGGTTTTGATGGCCACCAATCG 926
Db   780 CAATCAGCTGGATGGGGTTTGAGCCCACAAAAAAAACATCAAAATGGCTAGCAACCG 839

Qy   927 GATTGATTATCCTGGATCAAGGCCTCCTTAGGACCAGGACGGATAGACGGAAAATTGAATT 986
Db   840 ACTTGATATCCTGGATCCCGCTTTGTTGCCGCCGGATCGGATT 899

Qy   987 TCCAACCCCTAATGAAGAGTCTCGGCTGGATAGTAGTTACGAAGATGAAGATGAATGAA 1046
Db   900 TCCACCGTGGAAGCTCGTGCCAGCGGAGCGATCGGTGGGAGGCGGAGGGTACA 959

Qy 1047 TTTAATGCGTGGCATTGATTTGAAGAAGATTGCCGAGAAGGATGAATGGGAGCGAGGCATCTGGTGTGC 1106
Db   960 CCTGACGCGGAGGCATCAACTTGACGAAAGAGATCGATGCTCAGGAGC 1019

Qy 1107 TGAACTTAAGGCTGTTTGCACTGAAGCTGGAAAGTTGCTTTGAGGGAGGCGGAGGGTACA 1166
Db   [continued]

```
Db   1020  AGAACTTGAAGGGTGTGTTGCACCGAGGGAGCCAGGCATGTACCGGCTTCGGGAACGAGACGAGAGTGCA 1079
           ||||||||||  |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1167  CGTGACTCAGGAGGATTTTGAGATGGCCGTGGCGAAGGTCATGAAAAGGAGACTGAAA  1226
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1080  CGTCACGCAGGAGAGATTTGACCTAGCCACTGCCAAGATTCTCAACAAACACGATGACAA 1139
           ||||||||||||||||||||||||||||||||||||||||||||
Qy   1227  AAACATGTCATTGCGGAAGTTGTGTGGAAG 1254
           ||||||||||||||||||||||||||||||
Db   1140  GGAGGTTGCTGTGTCTCTCGAAACTGTTCAAG 1167
```

RESULT 4
US-10-128-714-7369
; Sequence 7369, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Broshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7369
; LENGTH: 1170
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-7369

Query Match   40.7%; Score 510.4; DB 14; Length 1170;
Best Local Similarity   65.4%; Pred. No. 2.1e-142;
Matches 764; Conservative 0; Mismatches 401; Indels 3; Gaps 1;

```
Qy     87  GGGCCTCCGCCACTACTATTCTCAACATCTCCACGACACGCTACTGGGTCAAATCG 146
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3  GGCGCTGACAACTACTATTCCCAATAAGATCGAGTATCAAAACTCGGATTATCCAGG  62
Qy    147  GACTCATAACCTCAACCGTCTCGGAGGCTCAGAGAAACGACCTCCAATTCTAGGGTGAGGAT 206
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db     63  CCAGGCTGTTTACGTCGCTGGAAGCACAGGCAACGACTATAATTCGGAGGTTCGGCCT 122
Qy    207  GCTGCGCGGAAGAATTACAGCTTCTGCGAGGAAACCCGGCTCTTATGTCGGTGAAGTTGTCAA 266
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    123  ACTACCAGAGAGGAGTTGGGGCTTCGGCTCAACACCTGGGTCTTATGTAGTGGGAGGTTGGTGAA 182
Qy    267  AGTAATGGGCAAGAACAAAGTCCTTCTTCAAGGTCCACCCGAGAGGAAAATATGTTGTTGGTGA 326
           ||||||||||||||||||||||||||||||||||||||||||||
Db    183  GGTGATGAGGCACCAAGAAGTTCTTGTCATGAGTTCCAGAAGGAAAAATATGTTCGGTTGGA 242
Qy    327  CATTGACAAAATATTGACATTACAAAGATTACTCCATCCACTACAGTTGCACTCCGCAA 386
           |||||||||||||||||||||||||||||||||||||||||||||
Db    243  CATTGCAGACGGGGTCGATTATTACCAGAAACTTACTGTGTGGGCAAGGGTGTTGCGGTTCGCTTTC 302
Qy    387  CGACAGTTATGTTCTTCACTTAGTTCTGCCAAGTAAAGTTGATCCAATGGTCAATCTGAT 446
           ||||||||||||||||||||||||||||||||||||||||||||
Db    303  AGACTCTTATAAAATTAGAAAAAAATGTTGCCATCGTCGTGGTTGGACCGGCTCGTTCGGCTTAT 362
```

```
Qy    447  GAAAGTTGAGAAAAGTTCCCGATTCTACATATGACATGATTGGTTGGTTTAGACCAGCAAAT 506
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db    363  GATGGTCGAGAGAGGTGCTGACAGCACATACAGGATAGATGATCGGAGCGCTTGATCACAAAT 422
Qy    507  TAAAGAAAATAAAAGAGGTCATTGAGCTTACCAATACGATCAAACATCCTGAGCCTGTTTGAAAGTCT 566
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db    423  CAAGGAGGAATCAAGGAAGTCATTGAGCTGGTCTCAAGCATCCAGRACTGTTCGGAGTCTCT 482
Qy    567  TCGCAACGCACAGCCCAAAAGGGTGTCTCCGCCTATGGGGCCACCTGGTACAGGTAAAAACATT 626
           ||||||||||||||||||||||||||||||||||||||||||||
Db    483  TGGTATCGCACRAGCCCAAAAGGTGTTCTTCCTTAACGGGCCCGGGAACCGGTAAAAACACT 542
Qy    627  GTTGGGCTRAGGGCAGTGGGTGCTCATCATCACTGACTGTACATTCATTCAGGGTGTCTGGTTCTGA 686
           ||||||||||||||||||||||||||||||||||||||||||||
Db    543  GCTCGCGCCCGAGCCGGGTAGCCCATCACACAGAGTTGCCGATTCATCCAGGGGTCAGGGGCTCGGA 602
Qy    687  GTTAGTTCAGRAAATACRATTGGAGAAGGTTCCTRGAATGGTCAGGGGAACTTTTTGTTATGGC 746
           |||||||||||||||||||||||||||||||||||||||||||||
Db    603  ACTRAGTCCAGAATACATTGGTGGAGGGTAGTCGCCATGGTGCCTGGTGTTCGTCATGGC 662
Qy    747  CAGGGAACATGCTCCATCAATTATCTTCATCGGATGAAAATTGACAGTATTGGGATCTGCTCG 806
           ||||||||||||||||||||||||||||||||||||||||||||
Db    663  TCGGAACAGTGCCAAGCATTATTCATCGACGAGAGATTGACAGTATCGGGTGTTCTAGCG 722
Qy    807  GATGGAAATCTGAAGTGGCAAGCGGTGATAGTGAGGGTACAGGGTACTATGCGTGGGAACTTCT 866
           |||||||||||||||||||||||||||||||||||||||||||||
Db    723  TATAGACTCGG---CTGGCTCGGCNAGTTCAGGACGGTTGCGCNGGGTCACAGTTGGAGCTGCT 779
Qy    867  CAACCAGTTGGGAGGGGATTGGAAGCTTCAAAATAAGATCAAAGGTTTTTGATGGCCAAATCG 926
           |||||||||||||||||||||||||||||||||||||||||||||
Db    780  CAATCAGCGTGGGATGGGTTTGAGCCCACAAAAAACATCAAAATCGGTAGGCTACGAACCG 839
Qy    927  GATTGATRATCCTGGGATCCAAGCCCTCCTTAGACCAGGACGGATRAGACCGGGAAAATTGAATT 986
           ||||||||||||||||||||||||||||||||||||||||||||
Db    840  ACTTGATAATCCTTGATCCCGCTTTGTTCGCGGCCCGGAAGGATTGACCGGAAGGATCGGAGTT 899
Qy    987  TCCAACCCCTAATGAAGAGTCTCGGCTGGGGGATRATTTTGAAAATCCATTCTAGAGAATGAA 1046
           ||||||||||||||||||||||||||||||||||||||||||||
Db    900  TCCACGCCCACCGGTGGAAGCTCGTGCCGGATRATTTCTACGCATTCACTCACGGTCGGATGAA 959
Qy   1047  TTTAATGCGTGGGCATTGATTTGAAGAGAATGGCGGACAGAAGATGAATGACGCATCTCGTGC 1106
           ||||||||||||||||||||||||||||||||||||||||||||
Db    960  CCTGGACGCGGAGGCAATCAACTTGACGAAGAGATTGCCAGRGAGAAGAATGGATGGCTCAGGAGC 1019
Qy   1107  TGAACTTAAGGCGTGTTTGCACTCGAAAGCTGTGGAATGTTTGCTTGACGGAGTCCGAGGGTACA 1166
           |||||||||||||||||||||||||||||||||||||||||||||
Db   1020  AGAACTGAAGGGTGTGTTGCGCACCGGAGGCAGGCATGTACCGGCTTCGGGGACGAGAGTGCA 1079
Qy   1167  CGTGACTCAGGAGGATTTTGAGATGGCCGTGGCGAAGGTGATGAAAAAGGGACTGAAA 1226
           |||||||||||||||||||||||||||||||||||||||||||
Db   1080  CGTCACGCAGGAGAGATTTGACCTAGCCACTGCCAAGATTCTCAACAAACACGATGACAA 1139
Qy   1227  AAACATGTCATTGCGGAAGTTGTGTGGAAG 1254
           ||||||||||||||||||||||||||||||
Db   1140  GGAGGTTGCTGTGTCTCTCGAAACTGTTCAAG 1167
```

RESULT 5
US-10-128-714-1369
; Sequence 1369, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Broshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23

; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; SOFTWARE: Patentin version 3.1
; NUMBER OF SEQ ID NOS: 8603
; SEQ ID NO 1369
; LENGTH: 1351
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-1369

Query Match   25.9%;   Score 324.2;   DB 14;   Length 1351;
Best Local Similarity   66.9%;   Pred. No. 2.3e-86;
Matches   477;   Conservative   0;   Mismatches 233;   Indels   3;   Gaps   1;

```
Qy  319  GTTGTTGACATTGACAAAAATATTGACATTACAAAGATTACTCCATCCACTAGAGTTGCA  378
Db  318  GTCGTTGACATTGCACGCGGGGTCGAATATTACGAAAACTTACTGTGGGCAAGCGTGTTGCG  377

Qy  379  CTCCGCAACGACACTTATGTTCTTCACTTAGTTCTGCCAAGTAAAGTTGATCCATTGGTC  438
Db  378  TTGCTTCAGACTCTTTATAAATTAGAAAAATGTTGCCATCGTCGGTTGACCCGCTCGTT  437

Qy  439  AATCTGATGAAAGTTGAGAAAGTTCCCGATTCTACATATGACACATGATTGGTTAGAC  498
Db  438  TCGCTTATGATGGTCGACAGCACATACGATATGACTGAGCGCCTTGAT  497

Qy  499  CAGCAAATTAAAGAAATAAAAGAGGTCATTCATTGAGCTACCAATCAAACATCCTGAGCGTGTTT  558
Db  498  CAACAAATCAACGAGGAGTCATTGAGCTTGGCTCTCAGAGCATCCAGAACTGTTC  557

Qy  559  GAAAGTCTTGGAAATTGCACAACCAAAGGGTGTCCTTGGCACCAGGGAACCGGT  618
Db  558  GAGTCTTCTTGGTATCGCACAGCCCAAAGGGTGTTCTTCTTTTACGGGGCCGCCCGGAACCGGT  617

Qy  619  AAAACATTGTTGGCTGAGGGCCAGTGGCCTCATCATTACTGACTGTGTTTCATTCAGGTGTCT  678
Db  618  AAAACACTGTCTGCGCCGAGCGGTAGCCATCACACAGATTGCCGATTCATCAGGGTCAGC  677

Qy  679  GGTTCTGAGTTAGTTCAGAAATACATTGGAGAAGGTTCTGAAATGGTCAGGGAACTTTTT  738
Db  678  GGCTCCGGAACTAGTCCAGAAAATACATTGGGTGAGGGGTAGTCCAGAGTCGCATGGTGTTC  737

Qy  739  GTTATGGCCAGGGAACATGCTCCATCAATTATCTTCATTGGAATGAAATTGACAGTATTGGA  798
Db  738  GTCATGGCTCGGAGACATGCACCAAGCATTAATCTTCATTGGAGACTGACATGCTGGT  797

Qy  799  TCTGCTCGGATGGAATCTGGAAGTGGCAACGGTAGTGAGGTACAGGGTACTATGCTG  858
Db  798  TCTAGCCGTATAGACTCGG---CTGGCTCTGGAGATTCAGAGCTCAAAATTGGAATGTTG  854

Qy  859  GAACTTCTCAACCAGTTGAATTGGAGCTTGAAGCTCATCAATTCATCAGGGTCAGC  918
Db  855  GAGCTCTCAATCAGCTGGATGGGTTGAGCCCACAAAAAACATCAAAATCATCATCATGGCT  914

Qy  919  ACCAATCGGATTGATTATCCTTGGATCTAAGCCCTCCTTAGACCAGATAGAACCGGAA  978
Db  915  ACGACCGACTTGATATCCTGGAATCCGGTCTTTGTTGGCGGCGAAGGATTGACCGGAAG  974

Qy  979  ATTGAATTTCCAACCCTAATGAAGAGTCTCGGGTGGATATTTTTGAAAAATCCA  1031
Db  975  ATCCGAGTTTCCACGCGCATCGTAAGATTCTGCCCTCTCTTCTTTTAACCCCA  1027
```

RESULT 6
US-10-128-714-6369

Qy
Db
Qy
Qy
Db
Db
Qy
Db
Qy
Qy
Db
Qy
Db
Qy
Qy
Db
Qy
Db
Db
Qy

Query Match   25.9%;   Score 324.2;   DB 14;   Length 1351;
Best Local Similarity   66.9%;   Pred. No. 2.3e-86;
Matches   477;   Conservative   0;   Mismatches 233;   Indels   3;   Gaps   1;

```
Qy  319  GTTGTTGACATTGACAAAAATATTGACATTACAAAGATTACTCCATCCACTAGAGTTGCA  378
Db  318  GTCGTTGACATTGCACGCGGGGTCGAATATTACGAAAACTTACTGTGGGCAAGCGTGTTGCG  377

Qy  379  CTCCGCAACGACACTTATGTTCTTCACTTAGTTCTGCCAAGTAAAGTTGATCCATTGGTC  438
Db  378  TTGCTTCAGACTCTTTATAAATTAGAAAAATGTTGCCATCGTCGGTTGACCCGCTCGTT  437

Qy  439  AATCTGATGAAAGTTGAGAAAGTTCCCGATTCTACATATGACACATGATTGGTTAGAC  498
Db  438  TCGCTTATGATGGTCGACAGCACATACGATATGACTGAGCGCCTTGAT  497

Qy  499  CAGCAAATTAAAGAAATAAAAGAGGTCATTCATTGAGCTACCAATCAAACATCCTGAGCGTGTTT  558
Db  498  CAACAAATCAACGAGGAGTCATTGAGCTTGGCTCTCAGAGCATCCAGAACTGTTC  557

Qy  559  GAAAGTCTTGGAAATTGCACAACCAAAGGGTGTCCTTGGCACCAGGGAACCGGT  618
Db  558  GAGTCTTCTTGGTATCGCACAGCCCAAAGGGTGTTCTTCTTTTACGGGGCCGCCCGGAACCGGT  617

Qy  619  AAAACATTGTTGGCTGAGGGCCAGTGGCCTCATCATTACTGACTGTGTTTCATTCAGGTGTCT  678
Db  618  AAAACACTGTCTGCGCCGAGCGGTAGCCATCACACAGATTGCCGATTCATCAGGGTCAGC  677

Qy  679  GGTTCTGAGTTAGTTCAGAAATACATTGGAGAAGGTTCTGAAATGGTCAGGGAACTTTTT  738
Db  678  GGCTCCGGAACTAGTCCAGAAAATACATTGGGTGAGGGGTAGTCCAGAGTCGCATGGTGTTC  737

Qy  739  GTTATGGCCAGGGAACATGCTCCATCAATTATCTTCATTGGAATGAAATTGACAGTATTGGA  798
Db  738  GTCATGGCTCGGAGACATGCACCAAGCATTAATCTTCATTGGAGACTGACATGCTGGT  797

Qy  799  TCTGCTCGGATGGAATCTGGAAGTGGCAACGGTAGTGAGGTACAGGGTACTATGCTG  858
Db  798  TCTAGCCGTATAGACTCGG---CTGGCTCTGGAGATTCAGAGCTCAAAATTGGAATGTTG  854

Qy  859  GAACTTCTCAACCAGTTGAATTGGAGCTTGAAGCTCATCAATTCATCAGGGTCAGC  918
```

RESULT 7
US-10-128-714-369

; Sequence 369, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 369
; LENGTH: 3350
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-369

Query Match   25.9%; Score 324.2; DB 14; Length 3350;
Best Local Similarity 66.9%; Pred. No. 4.1e-86;
Matches 477; Conservative 0; Mismatches 233; Indels 3; Gaps 1;

```
Qy   319 GTTGTTGACATTGACAAAAATATTGACATTACAAAGATTACTCCATCCACTAGAGTTGCA 378
             ||| |||||||||||||||| |||| |||||  ||||| || ||||||||||||||||
Db  1317 GTCGTTGACATTGCAGACGGGGGTCGATATTACGGAAACTTACTGTGGGCAAGCGTGTTGCG 1376

Qy   379 CTCCGCAACGACAGTATATTCTTCACTTAGTTCTGCCAAGTAAAGTTGATCCAATTGGTC 438
             ||||| |||| ||| || |||||||||| |||  |||||| || ||||||||||||
Db  1377 TTGCTTTCAGACTCTTATAAATTAGAAAAAATGTTGCCATCGTCGGTTGACCCGCTCGTT 1436

Qy   439 AATCTGATGAAAGTTGAGAAAGTTCCCGGATTCTACATATGACATATGACACAATTCCTG 498
             ||| ||||  ||||||||||||  ||||  |||||| ||| |||  |||||||| 
Db  1437 TCGCTTATGATGATGGTGGAGAAGGTGCCTGACAGCACAATAGTGATGTGGTTTAGAC 1496

Qy   499 CAGCAAATTAAAGAAATAAAAGAGGTCATTGAGCTACAAATCAAATCAAACATCCTGAGCTGTTT 558
             |||||||||  ||| || || |||| ||| ||||| ||| |||||||| ||||||
Db  1497 CAACAAATCACAAGAGATCAAGAAGTCATTGAGCTTGGCTCTCAAGCAATCCAGAACTGTTC 1556

Qy   559 GAAAGTCTTGGAATTGCACACCAAAGAGGGTGTCCGTCCTAGTGGGCCACCTGGTTACAGGT 618
             ||  |||| ||||| |||| |||| ||  | |||| |||||| ||| |||| |||
Db  1557 GAGTCTCTTGGTATCCGACAGCCCAAAGGTCTTCTTTTACGGGGCCGCCCGGGGAACCGGT 1616

Qy   619 AAAACATTGTTGGGCTAGGGCACTGGCCTCCATCATACTGAACTGTACATTCATCAGGGGTCT 678
             |||| | |||  |||  ||||   ||| || ||||| |||| | |||||| |||
Db  1617 AAAACACTGTCTCGCCCGGACCGGGGTAGCCCCATCACACAGATTGCCGATTCATCAGGGGTCAGC 1676
```

RESULT 8
US-10-128-714-5369

; Sequence 5369, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5369
; LENGTH: 3351
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-5369

Query Match   25.9%; Score 324.2; DB 14; Length 3351;
Best Local Similarity 66.9%; Pred. No. 4.1e-86;
Matches 477; Conservative 0; Mismatches 233; Indels 3; Gaps 1;

```
Qy   679 GGTTCTGAGTTAGTTCAGAAAATACATTGGAGAAGGTTCTAGAAATGGTCAGGGAACTTTT 738
             |||||||||| |||||| ||| |||| |||||||||||| ||||| |||||||||||
Db  1677 GGCTCGGAACTAGTCCAGAAATACATTGGGGTAGAATACATTGGGGTCGGCCATTGGA 1736

Qy   739 GTTATGGCCCAGGGAACATGCTCCATCAAATTATATCTCATGGATGAAATTGACAGTATTGGA 798
             |||||||| |||||||| |||| ||||| | ||||||||||||||||| ||||||||
Db  1737 GTCATGGGCTCGGAGAACGATGCACCAGAGCATTATTCTCATGGACGAGATTCGGGT 1796

Qy   799 ATTGAATTTCCAACCCCTAATGAAGAGTCTCGGGCTGGATATTTTGAAAATCCA 858
             |||  |||||||||||||||| | ||| ||| |||||| ||||| ||||||||
Db  1797 TCTAGCCGTATAGACTCGG---CTGGCTCTGGAGATTCAGAGGCTCAGATGTGG 1853

Qy   859 GAACTTCTCAACCAGTTGGATGGATTTGGAAGCTTCAAATATAAGATCAAAGGTTTTTGATCGGGCC 918
             ||||  ||||||| ||||||||| |||||||| |||| |||| |||||| ||||
Db  1854 GAGCTGCTCAAATCAGCTGGATGGGGTTGAGCCCCACCAGAGGACGGAATAGACCGGAA 1913

Qy   919 ACCAATCGGATTGATATCCTGGATCAACGGACGGGATAGACCGGAA 978
             |||||||| |||| ||||| |||| ||||||||| ||||||||||
Db  1914 ACGAACCGACTTGATATCCTGGAATCCGGTTGTTGCGGGCCGGAAGGATTGACCGGAAG 1973

Qy   979 ATTGAATTTCCAACCCCTAATGAAGAGTCTCGGGCTGGATATTTTGAAAATCCA 1031
             ||||  |||||||||||||||| | ||| ||| |||||| ||||| ||||||||
Db  1974 ATCGAGTTTCCACCGCCATCGTAAGATTCTGGCCCTTCTCTTCTTTTAACCCCA 2026
```

```
Db  1438  TCGCTTATGATGGTCCGAGAAGGTGCCTGACAGCACAATACGATATGATCGGAGGGCTTGAT  1497
                |||||      ||||    |  |||||   |   ||   ||  |||| 
Qy   499  CAGCAAATTAAAGAAATAAAAGAGGGTCATTGAGCTACCAATCAAACATCCTGAGCTGGTT   558

Db  1498  CAACAAAATCAAGGAGATCAAGGAAGTCATTGAGCTTCAACCATCAGAACTGTTC        1557
                ||||||  |  |||||||||        |||||  ||  |||| 
Qy   559  GAAAGTCTTGGAATTCCACAACCAAAGGTGTCCTGCCTCATGGGCCACCTGGGTACAGGT   618

Db  1558  GAGTCTTCTTGGTATGGCACACGCCCAAAAGGTGGTTCTTCTTTACGGCGGCCACGCGGT   1617
                ||| ||  ||       |||||
Qy   619  AAAACATTGTTGGCTGAGGGCAGTGGCTCATCATCATACTCAGGGTGTCT            678

Db  1618  AAAACACTGCTGCCCCCGAGCGGGTAGCCCATCACACAGATTGCCGAATTCATCAGGTCAGC  1677
                ||| ||||| 
Qy   679  GGTTCTGAGTTGGTTCAGAAATACACATTGGAAGAAGGTTCTAGAAATGGTCAGGGAACTTT  738

Db  1678  GGCTCGGAACTAGTCCAGAAAATACACATTGGTCGAGGGTAGTCGCATGGGTGCCTGACCT   1737
Qy   739  GTTATGGCCAGGGGACACATGCCTCATCAATTATCTCATGGATGAAATTGACAGTATTGGA   798

Db  1738  GTCATGGCTGCTGAGAACATGCACCAAGCATTATCTTCATGGACGAGATTGACAGTATCGGT   1797
Qy   799  TCCTGCTCGGATGGAAGTGGGCAACGGGTGGATAGTGAGGTACAGGGTACTATGCTCTG    858

Db  1798  TCTAGCCGTATAGACTCGG...CTGGCCTGGAGATTCAGAGTTGCAGCGGGTACAACAGTGTTG  1854
Qy   859  GAACTTCTCAACCAGTGGATGGATGGAGTTGAAGCTTCAAATAAGATCAAGGTTTTTGATGGCC   918

Db  1855  GAGCTGCTCAAATCAGCTGGATGGGTTGAGCCCACAAAAAAAATCATCATTCATGGCT   1914
Qy   919  ACCAATCGGATTGATATCCTGGGATCAACGCCCTCTTAGACCACGACGGGATAGAGACGGAA   978

Db  1915  ACGAACCGACTTGATATCCTGGAATCCCGCTTTGTTGCGGCCGCGAAGGATTGACCGGGAAG   1974
Qy   979  ATTGAATTTCCAACCCCCTAATGAAGAGTCTCGGGCTGGAATATTTGAAAATCCA      1031

Db  1975  ATCGAGTTTCCACCGCCATCGTAAGATCTCGCCCTTCTCTTCTTTTTAACCCCA        2027
```

Query Match 24.3%; Score 304.6; DB 10; Length 1332;
Best Local Similarity 61.7%; Pred. No. 1.8e-80;
Matches 503; Conservative 0; Mismatches 309; Indels 3; Gaps 1;

```
Qy   410  TTCTGCAAGTAAAGTTGATCCATTGGTCAATCTGATGAAAGTTGAGAAAGTTCCCGATT   469
```

```
Db   494  TTCTTCAAGATGAAGTGGATCCAATGGTGTCTGTGCTTGTGATGAAAGTGGAGAAGGCTCCTTTGG   553
Qy   470  CTACATAATGACATGATTGGTGGTTTAGACCAGCAAATTAAAGAAATAAAAGAGGTCATTG   529

Db   554  AGTCAGATGCTGACAATTCAGGAGGTTTAGAGCTCAGGATTCAGGGATTAAGGAAGCTGTTG   613
Qy   530  AGCTACCAATCAAACATCCTGAGCTGGTTGTTTGGAAGTTCTTGGAAATTCCACAACCAAAGGGTG   589

Db   554  AGTCAGATGCTGACAATTCAGGAGGTTTAGAGCTCAGGATTCAGGGATTAAGGAAGCTGTTG   613
Qy   590  TCCTGCCTCATGGGCCACCTGGGTACAGGTAAAACATTGTTGGCTGAGGGCAGTGGCTCATC   649

Db   674  TGATCTTTGATGGTGAGCCAGGCCACTGGGAAGAGCATTGCTTGCTAAGGCAGTCAGC   733
Qy   650  ATACTGACTGTACATTCATCAGGGTGTCTGGTTCTGAGTTGGTTCAGAAATACACATTGGAG   709

Db   734  CTACTTCAGCTACTTTCTTCCGTGTTGTCCGGTAGTGGAACTGATTGGAGAAGTATTTAGGAG   793
Qy   710  AAGGTTCTAGAAATGGTCAGGGAACTTTTGTTATGGCCAGGGACACATGCTCCATCAATTA   769

Db   794  ATGGTCCCAAGCTTCTGAGGGAACTTTCAGGGGTTGCTGAGTGACACTTGCTGATGAGGAC   853
Qy   770  TCTTCATCGATGAAATTGACAGTATTGGATCTTGTCTGCTCGGATGGAAATCTGGAAGTGG   829

Db   854  TTTTCATAGACGAGATTGATGCGTTGGGCACCAAGCCGGTATGA...CGCAAAACTCAGGTG   910
Qy   830  GTGATAGTGAGGTACAGCGTACTACTATGCTCGGAACTTCTCAACCAGTGGATGGATTGGAAG   889

Db   911  GGGGAAGGTGAAAATCCAAAGAACCATGTTTGAACTACTACTGAACCAGCTTGATGGGTTGAT   970
Qy   890  CTTCAAATAAGATCAAGGTTTTGATGGCCACCAATCGGATTGATATCCTGGATCAAGGCCC   949

Db   971  CAAGGAGGATGGTTAAGGTTATTCTCGGCGAAAACAGGATCGAGAGTCTTGACCCCGCAC   1030
Qy   950  TCCTTGACCAGGACGGACTAGACCGGAAAATTGAATTTCCACACCCCTAATGAAGAGTCTC   1009

Db  1031  TTCTTAGACCAGGCCGGACCTGACACAGAAAAATCGAAGATTTCCCTTCCAGAATATCAAAACAA   1090
Qy  1010  GGCTGGATAATTTGAAAATCCATTCTAGAAGAATGAAGTTTAATGCGGTGGGCCATTGATTGA   1069

Db  1091  GAAGACGTATCTTCCAGATACACACATCAAAGATGACTTTTGCTGAACATTTGACACTG   1150
Qy  1070  AGAAGATTGCCGAGAAGATGAAATGAGCATCTGGTGCTGAACTTAAGGCGTGTTGCACTG   1129

Db  1151  AACAGTTTGTTAATGACAAAAGACGGAGTTCAGGAGCTGGATAATATGACACAG   1210
Qy  1130  AAGCTGGAATGTTTGCTTTGAGGGGAGCCGAGGGGTACACGTGACTCAGGAGGAATTTTGAGA   1189

Db  1271  AGGCGAGAGGAGAAGGTTATGTTCAAGAAGAAGAA   1305
Qy  1190  TGGCCGTGCGGAAGGTGATTGAGGTGATTGAGAAAAAAGGAGAGACTGAA   1224

Db  1211  AAGCTGGTGTCTGCTTGCTCCTTAGGAGCCATGTGGACTTCAAGA   1270
Qy  1211  ...                                                1270
```

RESULT 11

Query Match　22.2%; Score 278; DB 12; Length 1566;
Best Local Similarity　56.4%;　Pred. No. 2e-72;
Matches 540;　Conservative　0; Mismatches 415; Indels 3; Gaps 1;

```
Qy  249 TGTCGGTGGAAGTTGTCAAAGTAAATGGGCCAGAACAACAAAGTCCTTGTGTCAAGGTCCACCCAGA 308
            ||| || || |||||| ||| | |||  |||| || || ||||| ||| | ||  |||||  |||||
Db  177 TGTCGGTGAGTCCTTAAGCAATTAACTGAGAGAAAAATTCATTGTTAAAGCTACAAATGG 236

Qy  309 AGGAAAAATATGTTGTTGACATTGACAAAAATATTGACATTACAAAAGATTACTCCATCCAC 368
            ||| |  | || ||  |||  ||| ||||||||| ||  | ||| ||  || ||||| ||
Db  237 ACCAAGATATGTTGGCGTTGTGCGTCCACACAGCTTGACAAAAGTAAAGCCAGGAAC 296

Qy  369 TAGAGTTGCACTCCGCAACGACAGCAGTTATGTTCTTCACTTAGTTCTGCCAGTTAAAGTTGA 428
            ||||||| |  |||   ||||||| |  |||   ||||||   | ||||| ||  ||||||
Db  297 AAGGAGTTGCCTTGGATATGAACTACACTAACTATCATGAGATATTTGGCCAAGACGCGGTGA 356

Qy  429 TCCATTGCTCAATCTGATGAAAAGTTGAGAAAGTTCCCGGATTCTTACATATGACATGATTGG 488
            ||| | | | ||  ||| | |||||| |||||| || ||||| |||| | | |||| |||||
Db  357 TCCACTTGTTTACACATCTCATGAGACACCCTGGGGAATGTTTCTTATTCTTGGAGATTGG 416

Qy  489 TGGTTAGACCAGCAAATTAAAGAAAATAAAAGAAGGTCATTGGAGCTACCAAATCAAAACATCC 548
            | || |||   ||||  || |||  |||  |||||  || || | | | ||| |||| ||||
Db  417 AGGTCCTGCAGAACAGATTCGGCGGAATTAAGAGGTAATAGTCTTTACAAATCC 476

Qy  549 TGAGCTGTTTGAAAGTTCTTGGAATTGCACAACCCAAAGGGGTGTCCCGCTCTTATCGGCCACC 608
            | ||||| |||||| |||| ||||||||| || |||   |||||| | | || || |||||
Db  477 AGAATTATTCCAGCGTGTAGGGAATAAATACCTCCAAAAGGCCTGTTTGTTATATCGGACCACC 536

Qy  609 TGGTACAGGTAAAACATTGTTGGCTAGGGCAGTGGCTCATCATACTGACTGTACATTCAT 668
            ||| |||  |||  |||| | |  |||  ||||||  ||||||  |||| ||||| |||
Db  537 AGGCACAGGAAAAACACTCTTGGCACGAGAGGTTCTAGAATGGTCAG 596

Qy  669 CAGGGTGTCGTGGTTCTGAGTTAGTTCAGAAATACATTGGAGAAGGTTCTAGAATGGTCAG 728
            ||||| || || ||||| ||||||| ||| | | |||| |||||||| |||||||||||
Db  597 AAAGGTGTATCGAGTTCTATTGTAGACAATGCATTTGGGTGAGTGCTCGTTCAGGTTAG 656

Qy  729 GGAACTTTTTGTTATGGCCAGGGAACATGCTCCCATCAAATTATCTTCATGGATGAAATTGA 788
            ||| ||||  |||| ||||||| | |||  || || || |||| ||  ||||| || |||||
Db  657 AGAAAGTTTAATTATGGCCAGGGACCATCAGCCTGATCATTTTATGGATGAAAATAGA 716

Qy  789 CAGTATTGGATCTGCTCGGATGGAAATCTGGAAGTGCAACGTGATCATTTTAGGATGAAAATTGA 848
            ||||| ||| || ||| ||||||||  | | ||| |  || ||| ||| | |||| || || |
Db  717 TGTTATTGGTGTTCGTCGGTTT---TCTGAAGGTACTTCAGCTGGAGGACGATTCAGAG 773

Qy  849 TACTATGCTGGAACTTCTCAACCAGTTGGGATGGATTTGAAGCTTCAAATAAGGATCAAGGT 908
            ||||| | |  || || || |  || ||  |||| ||||| | | |  | ||  | |||||
Db  774 AACCTTAATGGAGTTACTGAATCAAAATGGATGAATTGGATACTCTGCGATAGAGTTAAAAT 833
```

RESULT 12
US-10-032-585-6680
Sequence 6680, Application US/10032585
Publication No. US20030180953A1
GENERAL INFORMATION:
;   APPLICANT: Terry, Roemer D.
;   APPLICANT: Bo, Jiang
;   APPLICANT: Charles, Boone
;   APPLICANT: Howard, Bussey
;   TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
;   FILE REFERENCE: 10182-005-999
;   CURRENT APPLICATION NUMBER: US/10/032,585
;   CURRENT FILING DATE: 2001-12-20
;   NUMBER OF SEQ ID NOS: 8000
;   SOFTWARE: Patentin version 3.1
;   SEQ ID NO 6680
;   LENGTH: 1398
;   TYPE: DNA
;   ORGANISM: Candida albicans
US-10-032-585-6680

Query Match   21.9%; Score 275.2; DB 12; Length 1398;
Best Local Similarity 58.8%; Pred No. 1.3e-71;
Matches 494; Conservative 0; Mismatches 343; Indels 3; Gaps 1;

```
Qy    409 GTTCTGCCAAGTAAAGTTGATCCATTGGTCAATCTGATGAAGTTGAGAAAGTTCCCGGAT  468
              ||| |||||| ||||||  ||| ||| || ||  ||||| || ||||| |||  ||| |
Db    487 GTGTTACAAGATGATCCTGATCCCTANGGTTTCGGTTAATGAAACTGACAAATCCCTACA  546

Qy    469 TCTACATATGACATGATTGGTGGTTTAGACCAGCAAATTAAAGAAATAAAAGAGGTCATT  528
              ||| || |||||| | |||| || ||||| ||| || |||| |||  ||| || |||
Db    547 GAATCATCATCCCGACCATTGGGGGTTGGAATCACAAATCCAAAGAAATCAAGAATCAGTC  606

Qy    529 GAGCTACCAATCAAAACATCCTGAGCTGTTTGAAAAGTTCTTGGAATTGCACAACCAAAGGGT  588
              || |||| || ||  ||||| ||  || ||| ||| || |||||   |||||  ||| |
Db    607 GAATTGCCATTGACACATCCTGAATTGTACGGAGAAATCGGATCAAGCCCTCCAAAGGT  666

Qy    589 GTCCTGCTCTATGGGCCACCTGGTACAGGTAAAACATTCTTGGCTAGGGCAGTGGCTCAT  648
              || | || || ||||| ||| ||||| ||||||| |  |||| | | |||||| | |||
Db    667 GTGATATTATATGGGGCTCCCGGTACGGGTAAAACTTGTTGGCCAAAGCAGTGCCAAT  726

Qy    649 CATACTGACTGTACAATTCATCAGGGTGTCTGGTTCAGTTAGTTCAGAAAATCATTGGA  708
              ||  | || | || || ||| || || |  ||||| | ||||| |||| | ||| ||
Db    727 CAAACAAGTGCCACTTTCTTGAGAATGTCGGATCGGAGTTGATCCAGAAAAATATCTAGT  786

Qy    709 GAAGGTTCTAGAAGGTCAGGGAACTTTTGTTATGGGCCAGGGAACATGTCTCCATCAATT  768
              || | ||| | |||| ||  ||| ||| | |||| |||  | |||| |||||| |||||
Db    787 GATGGGTCCAAGGTTGTGTCGTCAAATTTCCAAATGCTCGTGATCACCGCACTTCAATT  846

Qy    769 ATCTTCATGGATGAAAATTGACAGTATTGGAATCGCTCGGATGGAATCTGGAAGTGGCAAC  828
              ||| ||||| |||||  |||| ||| | |||||||||  | |||||||| ||||||| 
Db    847 GTTTCATTGATGAAATCGAATCGGAAGAGAATACGAGTCTACATCTGG---T  903

Qy    829 GGTGGATAGTGGAGGTAACGGTACTATGCGTGGGAACTTCTCACCAGTTGGGATTGGAA  888
              ||||| | ||| |||| | ||||| ||  | ||  ||| ||| ||||||| ||| |||
Db    904 GGTGGAACGTGAAATCCAAAGAACCATGTTGGGAGTATTGAATCAGTTGGAGATGGGTTCGGAT  963

Qy    889 GCTTCAAATAAGATCAACAGGTTTTGATGGCCACCAATCGGATTGATATCCTGGATCAAGCC  948
              || ||||||| |||| | | ||||||||| ||| | | || | | ||||||| |||  ||
Db    964 GATAGAGGTGACATCAAAGTTATTATGGCTACAAAATAAGATAAGAGATCACTTGATCCTGCA  1023

Qy    949 CTCCTTAGACCAGGAACGGATAGAGCGTCCGCGAAAAATTGAATTTCAACCCCTAATGAAGTCT  1008
              || ||  | || || | ||| |||| |||  | |||| |  ||||| |||  ||| |  ||
Db   1024 TTAATCAGACCTCGGAAGAATCGATTAGAAAGGAATGATTTTGTTTGAAAACCCTGACCGCAATACC  1083

Qy   1009 CGGCTGGATATTTTGAAAAATCCATTCTTAGAAGAAGAATGATTTTGTTTGGGGCCATTGATTTG  1068
              ||||| || || ||||||| |  || ||| ||||| |||| ||| || | |||||  |||| |
Db   1084 AAGAAAAAGATCTTGACTATTTCACACCTCAAAGATAGTAGTTTGGCAGATGACTTAACTTG  1143

Qy   1069 AAGAAGATTGCCGAAGAAGATGAAATGGAGCATCTGGTGGCTGAACTTAAGGCTGTTGCACT  1128
              ||||| ||||| ||||| ||| |||| ||| |||| | || ||| ||||| ||| | |||
Db   1144 GATGAAATAGTCACTGGTAAGGATGAATTGTCCGGGTCGTGCTGAATATCAAAGCTATATGTACA  1203

Qy   1129 GAAGCTGGAAATGTTTGCTTTGAGGGAGCGGCGGAGGGGTACACGTGACTCACGGAGGATTTTGAG  1188
              |||| | || | |||| | || || ||| ||  | | |||| ||  | | ||| | ||||| |
Db   1204 GAAGCAGGGGTTATTAGCATTAAGAGAATAGGAAAGACGTATGCAAAGCTGAAGAATTTCAAA  1263

Qy   1189 ATGGCCGTGGCGAAGGTGATGAAAAAGGAGACTGAAAAAAAACATGTCATTGCGGAAGTTG  1248
              ||| | |||| | || || ||  ||| |||| ||||||| | |  | ||| ||  |||||
Db   1264 TCTGCAAAGGAAAAGAGTCTTGAAAAATAAGGTTGAAGAAAAACCTCGAAGGTTTGAATTG  1323
```

RESULT 13
US-10-101-510-379
Sequence 379, Application US/10101510
Publication No. US20030148295A1
GENERAL INFORMATION:
;   APPLICANT: WAN, JACKSON
;   APPLICANT: WANG, YIXIN
;   TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
;   FILE REFERENCE: 15117.0012
;   CURRENT APPLICATION NUMBER: US/10/101,510
;   CURRENT FILING DATE: 2002-03-20
;   PRIOR APPLICATION NUMBER: 60/276,947
;   PRIOR FILING DATE: 2001-03-20
;   NUMBER OF SEQ ID NOS: 805
;   SOFTWARE: Patentin Ver. 2.1
;   SEQ ID NO 379
;   LENGTH: 1891
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-101-510-379

Query Match   21.7%; Score 272.2; DB 12; Length 1891;
Best Local Similarity 56.0%; Pred No. 1.2e-70;
Matches 536; Conservative 0; Mismatches 418; Indels 3; Gaps 1;

```
Qy    250 GTCGGTGAAGTTGTCAAAGTAATGGGCAAGAACAAAGTCCTTGTCAAGGTCCACCCAGAA  309
              || |||||||| | |||  |  | ||| |||   ||  | ||||| |||| |||| |
Db    328 GTGGGTGAAGTGGCTTAAACAGTTAACTGAAGAAAATTCATTGTTAAAGCTACCAATGGA  387

Qy    310 GGAAAATATGTTGTTGACATTGACAAAAATATTGACAATTACAAAAGATTACTCCATCCACT  369
              || |||||||| ||||| || | |||| |||| | | |||||  | | |||| |||| |
Db    388 CCAAGATATGTTGTGGGTTGTCGTCGCAGCGTTGACAAAAGTAAGCCAGGAACA  447

Qy    370 AGAGTTGCACTCCGCACAGCACAGTATGTTGTTCACTTAGTTCTGCCAAGTAAAGTTGAT  429
              ||||||||  | || || | | |||| |||| || |||| |||| ||||||  |||||
Db    448 AGAGTTGCTTTGGAATATGACTACAACTAACTATCATGAGATATTGCCGGAGACAGGTCGGAT  507

Qy    430 CCATTGGTTCAATCTGATGAAAGTTGAGAAAGTTCCCGGATTCACATATGACATGATTGGT  489
              ||| ||||| ||| ||||||| ||||| |||| ||| |||| ||| ||||| | | |||
Db    508 CCACTGGTTTATAACATGTCTCATGAGGACCCCTGGGGAAATGTTTCTTATTCTCGAGATTGGA  567

Qy    490 GGTTTAGACCAGGAAATTAAAGAAATAAAAGAGGTCATTGAGCTACCAATCAAAACATCCT  549
              ||| | ||||| || || |||||| | |||| | |  ||||| ||| ||| |||||| |
Db    508 CCACTGGTTTATAACATGTCTCATGAGGACCCCTGGGGAAATGTTTCTTATTCTCGAGATTGGA  567
```

```
;    NUMBER OF SEQ ID NOS: 8603
;    SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2522
;    LENGTH: 1320
;    TYPE: DNA
;    ORGANISM: Aspergillus fumigatus
US-10-128-714-2522

  Query Match       21.7%;  Score 271.8;  DB 14;  Length 1320;
  Best Local Similarity  55.1%;  Pred. No. 1.1e-70;
  Matches  553;  Conservative   0;  Mismatches  447;  Indels   3;  Gaps   1;
```

GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Hu, Wenqi
APPLICANT: Tishkoff, Daniel
APPLICANT: Zamudio, Carlos
APPLICANT: Broshkin, Alexey M
APPLICANT: Lemieux, Sebastien M
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
TITLE OF INVENTION: Methods of Use
FILE REFERENCE: 10182-018-999
CURRENT APPLICATION NUMBER: US/10/128,714
CURRENT FILING DATE: 2002-04-23
PRIOR APPLICATION NUMBER: US 60/285,697
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: US 60/287,066
PRIOR FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: US 60/295,890
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/303,899
PRIOR FILING DATE: 2001-07-09
PRIOR APPLICATION NUMBER: US 60/316,362
PRIOR FILING DATE: 2001-08-31

RESULT 14
US-10-128-714-2522
; Sequence 2522, Application US/10128714
; Publication No. US20030119013A1

```
Qy    209 TGCGCGAAGAATTACAGCGTTCTCGCAGGAACCCGGCGTCCTTATCGTTGGTTGAAGTTGTCAAAG 268
          || |      || ||  ||  | ||    |||   ||||| ||| |   |  |   |     |  |
Db    272 TGGCGGAGGAACAGCCATTGCAGGTGGCTCGCTGTACAAAGAGATCATTTCCGACGAGAAAG 331

Qy    269 TAATGGGCAAGAACAAAGTCCTTGTCAAGGTCCACCAGAGAGGAGAAAATATGGTTGTTGACA 328
          || |    | |  | |   | ||    |       | | |    | |||| | | |   |
Db    332 ATCCCGATAAGAGCAAAATATGTCATCAAAGCTCAAAGCAGAATCGCCAAGTTTGTTGGTGAATT 391

Qy    329 TTGACAAAAATATTGACATTACAAAAGATTACTCCATCCACTAGAGTTGGCACTCCGCAACG 388
          ||  | | | | | | | || || |  || ||  | | | | |  |  | | | |   |  |
Db    392 TAGGGAGGAGAGTGAGCCCGACAGGGCATATCGAAGAAGGCATGAGAGTCGGCGTTCGACCGGA 451

Qy    389 ACAGTTATGTTCTTCACTTTAGTTTCTGCCAAGTAAAGTTGATCCATTGGTCAATCTGATGA 448
          | |||   | || |  ||| || |  | | || |  |  | | | |   || | | | ||
Db    452 ATAAAATACCAGATCATGTTTACCCCTTCCTCCTTAAGATCGAACCGTGTACGGGCAAGACCCTCT 511

Qy    449 AAGTTGAGAAAGTTCCCGAATTCTCACATATGACATGAATTGGTGTTTAGACCAGCAAATTA 508
          |  |||| |   | |  |  |  | | | | | | | |   | | ||| |  |  |  | |
Db    512 CGGTCGAAGACAAGCCCGACGTGACCTACGGCGGATGTTGGTGGATGCAAGGAGCAGATCG 571

Qy    509 AAGAAAATAAAAGAGGTCAATGAGCTACCAATCAACAATCCTGAGCTCGTTGAAAGTCTTG 568
          | | | || |  |  |   || |  |  ||    |   ||   |   |  ||| | |
Db    572 AGAAGCTGAGAGAAGTTGTCGGAAAATGCCTTTGCGTGTCACCGGAACGAGATTCGTCAACCTCG 631

Qy    569 GAATTGCACAACCAAAGGGGTCCTCGCTCCTATGGGCCACCTGGTACAGGTAAAACATTGT 628
          || || | |   |  |   | ||    | ||   ||| |   ||  |  |  | | |  |
Db    632 GTATCGGATCCGCCACAAGGGGCGCGGTTGAGGCTGCTATCTCCCCGGGTACGCAAGAGCCCTCT 691

Qy    629 TGGCTAGGGCAGTGGCCTCATCATTACTGACACTGTTGAGATTAAGAATTAAGGTAGTTTTGA 688
          |  || || | |   || |  | | | |  | || ||       | | |     | |  |  |
Db    692 GCGCTCGAGCAGTCGCTAACCGGACGGGACCCACTTTCATCCGTGTTATTGGGTTCTGGT 751

Qy    689 TGGTTCAGAAATACATTGGAGAAGAGGTTCTAGAAATGGTCAGGGAACTTTTGTTATGGCCA 748
          ||||  |  | ||  | ||  | |   ||   |  |  |  | || | |  ||| |  ||||
Db    752 TGGTTCAAAGTACGTTGGTGAGGGTCCCACGGAATGGTTCGAGAACACTTTTTGTTTTGGCCC 811

Qy    749 GGGAACATGCTCCATCAATTATCTTCATGGATGAAAATTGACAGTGATTGGATCTGCTCGGA 808
          ||| || |  | ||   | || |     | | | | | ||   ||  |  ||||  | |||
Db    812 GTACCAAGAAGGGGTGCATCACCTTCTTTGAGAAATGCCGGGTGGGCGGGTGCTCGGTT 871

Qy    809 TGGAATCTGGAAGTGGCAACGGTGATAGTGAGGGTACAGCGTACTATGCTGGAACTTCTCA 868
          || || ||||   || |   | | || | || |||| || ||   ||| | | |  | | |
Db    872 TCGATGAATGGAGCTGG---TGGTGACAACGAGGTTCAGGGTACTACTGCTTGAGTTGGATTA 928

Qy    869 ACCAGTTGGATGGAATTCAAATAAGATCAAGGTTTTGAATGGCCACCAATCGGA 928
          | ||||  |   | |   | | |  |  |  | | | ||  | | |||  |||||
Db    929 CACAGCTCGGATGTGTTTGATGCCCGTGTAACATTAAGGTCATTGTGCAACCGAC 988

Qy    929 TTGATATCCTGGGATCAAGCCCTCCTTAGACCAGGACGGATAGAACGGGAAAATTGAATTC 988
          ||| | |   ||| ||  |   | || ||    | | ||   || || | ||| || |
Db    989 CGTCGACCTTGGATCCTCGCCCTTGATGGCTCCTGGGTCGGTAGAGAATGAAGA 1048

Qy    989 CAACCCCTAATGAAGAGACTCTCCGGCGTGGATATTTTTGAAAAATCCATTCTTAGAGAGAT 1048
          | | | | | | | |  | || ||  |  |  |  |   | | | |  | |  | ||| ||
Db    1049 CGCTTCCCGAGTGGAAGGTCGGCCCAACATCCTCCCGCATTCATGCCAGAGAGTATGTCAG 1108

Qy    1049 TAATGCGTGGCATTGATTTGAAGAAGATTGCCGAGAAGAGATGGAGCAGCACTCTGGTTGCTG 1108
           || | | |  | | | | |  | | | ||  ||| |||  |  | | |  ||  |   | | | |
Db    1109 TCGAAAGAGAGACATCCGGTGGGAGCTGAATCTCCGACTATGTCCTAATGCTACCGGTGCGG 1168

Qy    1109 AACTTAAGGCTGTTTGCACTGAAGCTGGAATGTTTGCTTTGAGGAGGACGAGCGGTACACG 1168
          |   | |||   |  | ||      ||  |  | | | | | | |
```

```
Db   1169 AACTACGGAGAGTGTTGCCACAGAAGCTCGGGGGGCGGGAAGGTTG
          ||||   ||| ||       ||||     |||||   ||||   |||||
Qy   1169 TGACTCAGGAGGATTTTGAGATGGCCGCTGGGGGCCGAAGGTTGATGAA 1211

Db   1229 CCACAGAGAAGGACTTCCTCGGCTGCTGTGGACAAGGTCATCAA 1271
          ||   |||||||||||||||||||||||||||||||||||||||
Qy   1169
```

RESULT 15

Query Match   21.7%;   Score 271.8;   DB 14;   Length 1320;
Best Local Similarity   55.1%;   Pred. No. 1.3e-70;
Matches   553;   Conservative   0;   Mismatches   447;   Indels   3;   Gaps   1;

```
Qy    209 TGGCGAAGAAATTACAGCTTCTGCAGGAACCCGGCTCTTAATGTCGGTGAAGTTGTCAAG 268
          ||||   ||| ||||  |||||   |||||  ||| |||||  ||  ||||||| |||||
Db    272 TGGCGGAGGAGAACAGCCATTGCAGGTGGCTCGCGTCTGTACAAAGATCATTTCGACGAGAAG 331

Qy    269 TAATGGGCAAGAACAAAAGTTCCTTGTCTTCAAGGTCACACCCAGAAGAGAAAAATAGTTGTTGACA 328
          ||  |||||   |||||||    ||  |||| ||||||  ||||||  ||||| ||  ||||||
Db    332 ATCCCGATAAGAGCAAAATATGTCATCAACGTCAAGCAACAGAATCCCAAGTTGTCGTGAATT 391

Qy    329 TTGACAAAAATAATTGACATTACAAAGATTACTCCATCCACTAGAGTTGCACTCCGCAACG 388
          ||||  |||   |||| | ||   |||||| |  |||||| || |||||||||||||||||
Db    392 TAGGGAGAGAGTGAGCCCGACCGGATATCGAAGAGAGAAAGGCATGAGAGTCGGCGTTGACCCGGA 451

Qy    389 ACAGTTATGTTCTTCACTTAGTTGTTGCCAAGGTAAAGTTGATCCAATCTGATGA 448
          ||||| |||| | ||||||||   |||   |||  |||||||| |||||||||||
Db    452 ATAAATACCAGATCATGTTTACCCCTTCCTCCTAGATCGGACCCTAGCGTGACGATGATGA 511

Qy    449 AAGTTGAGAAAGTTCCCGATTCTACATATTGACATGATTGGTGGTTTTAGACCACCAAAATTA 508
          |||||| ||| |||   || || || || |||| |||| ||| |||| |||| |||||||
Db    512 CGGTCGAGAACAAGCCGGACGTGACCTACGGCGATGGTGGCGATGCAGAGGAGCAGAATCG 571

Qy    509 AAGAAAATAAAGAGAGGTCATTGAGCTACCAAATCAAACAATCCTCGAGGCTGTTTGGAAAGTTCTTG 568
          ||||| || ||||| |||| || | |||||   || ||||||| |||||   |||| ||| ||
Db    572 AGAAGCTGAGAGAAGTGTCGAAATGCCTTGCTGTCACCGGAACGATTCGTCAGCTAAAAACATTGT 631

Qy    569 GAATTGCACCACCAAAGGGTGTCCTGCTCCTATGGGCTACAGGTAAAACATTCT 628
          || |||||  || |||| |||||| |||||| |||||| ||||  ||||||||
Db    632 GTATCGATCCGCCCAAGGGCGGGGCCGGGCTGCTGTACGGGTCCCCGGTACCGGCAGAGACCCTCT 691

Qy    629 TGGCTAGGGCGCAGTGGCTCATCATACTGACTGTACATTCATCAGGGGTGTCTGGTTCTCGAGT 688
          |||||  |||| |||||||||   ||| || |||| | |||||||| |||||| |||| |||||
Db    692 GCGCTCGAGCAGTCGCTTAACCGACGAGCCGCACTTTCATCCGTGTTATTGTAGCGAGT 751

Qy    689 AAGTTGAGAAATACAATCATTGGAGGAAGAGTTCTAGAATGGTTCAGGGGAGAAACTTTTTGTTATGGCCA 748
          ||    || |  |||| | |||  | ||  ||||  |||||||||||| || ||| |||| ||||
Db    752 TGGTTCAAAAGTTACGTTGGTGCAGAAATGGTTCGGAGAAACTTTTTGGAGATGGCCC 811

Qy    749 GGGAAACATGCCTCCATCAAATTAATCTTCATGGATGAAAATTGACAGTATTGGGATCTTGCTCGGA 808
          |||  ||  |  |  |||  ||| || || |||||||| || ||| |||||||| || || |||||
Db    812 GTACCAAGAAGGCGTGCCATCATCTTCTTGACGAAAATCGAACGGGTGTCTGCTCGTT 871

Qy    809 TGGAATCTGGAAGTGGCAAACCGGTGATAGTGAGGTACAGCGCTACTATGCTGGAACTTCTCA 868
          ||||||||||| |||||| |  ||||  |||| ||||| |||| || ||||||||||||||||
Db    872 TCGATGATGAGCCTGG---TGGTGACAACGAGGTTCAGCCGTACTATGCTTGAGTTGATTA 928

Qy    869 ACCAGTTGGATGGATTGAAGCTTCAAATAAGATCAAAGGTTTTGATGGCCACCAATCGGA 928
          ||||||||| ||||||||| ||| |||| ||| || |||||| || |||||||||||||
Db    929 CACGAGCTGGGATGGTTTTGATCCGGTAACATTAAGGTCATGTTTGCAACCAACGGAC 988

Qy    929 TTGATATCCTGGATCAAAGCCCTCCTTAGACCAGGACGTAGACCGGAAAAATTGAATTTC 988
          ||||||| ||||| |||  ||| ||    | ||||||||||  ||||||  |||||||||
Db    989 CGTCGGACCTTGGATCCCGTGCCCTGATCGCGTCGGTGTTGGACCGCAAGATCGAATCT 1048

Qy    989 CAACCCCTAATGAAGAGAGTCTCGGCTGGATATATTTGAAAAATCCAATTCTAGAGAAGAATGAATT 1048
          || ||| || |||||| | |||||| ||   |||| |||||||| || || |||| |||| |||||
Db   1049 CGCTTCCCGATGTGGAAGGTCGCGCCCAACATCCTCCGCATTCATGCCAAGAGTATGTCAG 1108

Qy   1049 TAATGCGGTGGCATTGAATTTGGAAGAAGATTGCCGAGAAGAATGAAATGGAGCATCTGGTGTCTG 1108
          ||| | |||   ||| || ||||| |||||| | |||   |||| || |||| ||| |||
Db   1109 TCGAAAGAGACATCCGGTGGGAGCTGAATCTCCCGAACTATGTCCTAATGCTACCGGTGCCG 1168

Qy   1109 AACTTAAGCGTCGTTTGCACTGAAAGCTGGAATGTTTGCTTTGAGTGGATGAATGAA 1168
          | ||| ||  |  ||||| |   ||||| || |||  ||| || |||| |||||||
Db   1169 TGACTCAGGAGGATTTTGAGATGGCCGCTGGGGGCCGAAGGTTGATGAA 1211

Qy   1169 AACTACGGAGAGTGTTGCCACAGAAGCTCGGGGGGCGGGAAGGTTG 1228
          ||||   ||| ||       ||||     |||||   ||||   |||||
Db   1229 CCACAGAGAAGGACTTCCTCGGCTGCTGTGGACAAGGTCATCAA 1271
```

Search completed: November 9, 2003, 12:44:10
Job time : 479 secs

QM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.6

Run on: November 9, 2003, 08:53:50 ; Search time 3025 seconds
                                            (without alignments)
                                            10075.310 Million cell updates/sec

Title: US-09-462-972-1
Perfect score: 1254
Sequence: 1 atggcttctgtaggagtcga.........cattgcggaagtgtggaag 1254
Scoring table: IDENTITY NUC
               Gapop 10.0 , Gapext 1.0
Searched: 22781392 seqs, 12152238056 residues
Total number of hits satisfying chosen parameters: 45562784
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   EST:*
               1: em_estba:*
               2: em_esthum:*
               3: em_estin:*
               4: em_estmu:*
               5: em_estov:*
               6: em_estpl:*
               7: em_estro:*
               8: em_htc:*
               9: gb_est1:*
               10: gb_est2:*
               11: gb_htc:*
               12: gb_est3:*
               13: gb_est4:*
               14: gb_ests:*
               15: em_estfun:*
               16: em_estom:*
               17: em_estfun:*
               18: em_gss_inv:*
               19: em_gss_pln:*
               20: em_gss_vrt:*
               21: em_gss_fun:*
               22: em_gss_mam:*
               23: em_gss_mus:*
               24: em_gss_pro:*
               25: em_gss_rod:*
               26: em_gss_phg:*
               27: em_gss_vrl:*
               28: gb_gss1:*
               29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	806	64.3	1655	11	AY103724	AY103724 Zea mays
2	664.6	53.0	858	10	BG645412	BG645412 EST507031
3	632	50.4	723	9	AJ497641	AJ497641 AJ497641
4	630.2	50.3	1302	11	AK010505	AK010505 Mus muscu
5	622.6	49.6	1649	11	BC045221	BC045221 Xenopus 1
6	614	49.0	742	14	CA784406	CA784406 sat98h03.
7 c	608.4	48.5	794	12	BI968213	BI968213 GM830004B
8	608.4	48.5	803	14	CB979660	CB979660 CAB70001
9	605	48.2	685	10	BE942284	BE942284 EST421865
10	604	48.1	1313	11	AK012294	AK012294 Mus muscu
11	602.8	48.1	715	12	BI311279	BI311279 EST531302
12	591	47.1	591	12	BM308101	BM308101 sak41g08.
13 c	587.8	46.9	699	12	BI945194	BI945194 sb3ell.y
14	587.6	46.9	627	12	BI944357	BI944357 sa61b06.
15	580.8	46.3	632	9	AW458284	AW458284 sh80g08.y
16	577.8	46.1	948	10	BE036136	BE036136 MO20d12 M
17	569.6	45.4	597	10	BM555495	BM555495 sp90a11.y
18	569.6	45.4	597	12	BM519523	BM519523 sak79a11
19	563.6	44.9	976	10	BE416520	BE416520 MUG011.F0
20	562	44.8	1162	14	CD506927	CD506927 CDA82-A11
21	560.2	44.7	618	9	AW757204	AW757204 s130e03.y
22	560	44.7	560	12	BM731600	BM731600 sa18lg12.
23	557	44.4	569	12	BI942538	BI942538 si96b05.y
24	555	44.3	589	12	BM309454	BM309454 sak58h07
25	554.8	44.2	583	12	BI674283	BI674283 sa09e06.
26	550	43.9	565	12	BM526265	BM526265 sal39a05.
27	547	43.6	547	12	BM527342	BM527342 sa160g01.
28	546.4	43.6	572	12	BI674291	BI674291 saq11b11.
29 c	542.8	43.3	1201	9	ALS39095	ALS39095 saqimb11.
30	542.8	43.2	617	13	BQ610324	BQ610324 sap0gc03.
31	541.4	43.2	543	12	BI944393	BI944393 saa90f03.
32	541.4	43.2	547	12	BM085422	BM085422 saj36e12.
33	540.6	43.1	563	12	BI699091	BI699091 sag22e12.
34	537.8	42.9	837	14	CA807968	CA807968 CA12Lt1O4
35	537.4	42.9	603	9	AW394562	AW394562 sh32e12.Y
36 c	533	42.5	638	12	BI418881	BI418881 LjNEST36h
37	532	42.4	1166	14	CD508516	CD508516 CDA91-C08
38	529.2	42.2	932	10	BE416524	BE416524 MUG011.F0
39	528.8	42.2	733	12	BI920689	BI920689 EST540624
40	526.4	42.0	697	10	BF459871	BF459871 067F11 Ma
41	525	41.9	907	10	BE416748	BE416748 MUG012.G1
42	523.4	41.7	553	12	BI944282	BI944282 saa07b04.
43	523	41.7	755	10	BG590658	BG590658 EST498500
44	520	41.5	559	12	BM885596	BM885596 samo2h03.
45	516.8	41.2	896	10	BE416483	BE416483 MUG011.CO

ALIGNMENTS

RESULT 1
AY103724

LOCUS   AY103724   1655 bp   mRNA   linear   HTC 16-OCT-2002
DEFINITION Zea mays PCO078380 mRNA sequence.
ACCESSION   AY103724
VERSION   AY103724.1   GI:21206802
KEYWORDS   HTC.
SOURCE   Zea mays
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 1655)
AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
          Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE   Maize Mapping Project/DuPont Consensus Sequences for Design of
          Overgo Probes
JOURNAL   Unpublished (2002)
REFERENCE   2 (bases 1 to 1655)
AUTHORS   Coe,E.H
TITLE   Direct Submission
JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
          Missouri, Columbia, MO 65211, USA
COMMENT   If you are interested in getting corresponding physical clones,
          these are publicly available from ZmDB and may be found by BLAST
          searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
          www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the

```
FEATURES
     source          1..1655
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:634343"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Consensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
```

maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

```
BASE COUNT       463 a    359 c    438 g    395 t
ORIGIN
```

Query Match   64.3%;   Score 806;   DB 11;   Length 1655;
Best Local Similarity   80.4%;   Pred. No. 2.5e-203;
Matches 944;   Conservative   0;   Mismatches 230;   Indels   0;   Gaps   0;

```
Qy    81  GGGCGAAGGGCCCTCCGCCACTACTATTCTTCAAACATCCACGAGCATCCACGTTCCTTCTTCG  140
          | |||  |||||||| ||||  |      |     |||| |||| ||||||||||||| ||| | ||
Db   126  GGGGGAGGGGGCTGGGGAGAATCAAAAGGTTCTCGGGTAACAAAGGTTCTCGTGAAGGTTCCACT  185

Qy   141  CCAAAAGACTCATAACCTCAACCGTCTCGGAGGCTCAGAGAAACGACCCTCAATTCTAGGGT    200
          ||||||| | |||   | || || ||||||||||||||||||||||||| |||||||| ||||
Db   186  GCAGAAGACCCATAACCTCAACCGCCTCGGAGGCCCAAGCGAACGACCCTCAACTCCCGAGT    245

Qy   201  GAGGGATGCTGCCGGAGAATTACAGACTTCTGCAGGAACCCGGCTCTTTATGTCGGTGAAGT    260
          ||||  |  ||| ||||   |||||| |||||  |||||||||||||  | ||||||||||||
Db   246  TAGAATTGCTCAGGGAAGCAGTTGGCAGTTGCTTCAAGACCCTGGCTCATAGTGTTGGTGAGGT    305

Qy   261  TGTCAAAGTAATGGGGAAGGAACAAAAGTCCACTGCCCAGGAAGGAAAATATGT             320
          ||| ||||  ||  |||| ||||| ||  || |  |||||| |||||| |||||
Db   306  GGTGAAGGTCATGGGGAAATCAAAAGGTTCTCGTGAAGGTACATCCCGAGAGGTCAAATATGT    365

Qy   321  TGTTGACATTGACAAAAATATTGACATTACAAAGATTACTCCATCCACTAGAGTTCACT      380
          |||||| |||||||  |   |  ||  |  ||| ||| || ||||||  |||||||||||
Db   366  GGTGGATTAGATAAGAGGCATTGATATCACTAGGATCACACCTTCACACAAGAGTTGCTCT    425

Qy   381  TCCGGAATGACAGTTATATGCTCCATCTCGATCCTACCCAGCAAAGTTGATCCATTGGTCAA     440
          ||||| || || | |  |  ||||| | ||| ||||||| |||||| |||  ||| ||||||
Db   426  CCCCAACCACACTTTATTTCTTCACTTAGTTCTGCCAAGTAAAGTTGATCCATTCATTGGTCAA    485

Qy   441  TCTGATGAAAGTTGAGAAAGTTCCCGGATTCTACATATGACATGATTGGTGGTTTAGACCA     500
          |||| || ||| | |||  ||| ||||||| || |||| || ||| | |||||| |||||
Db   486  TCTCATGAAAGTTGTGAAGAGGTTCCGGATTCTACCTATGATTGATTGGTGAGGCCCTGAACCA    545

Qy   501  GCAAAATTAAAGAAAATAAAAAGAGAGTCATTGAGGCTAACAAAACATTCCTCGAGCTCGTTTGA    560
          || ||||| || |   |   | || || |  ||| || || |||||||| |||||||| |||||
Db   546  GCAAATTAAAGAGATCAAAAGAGGTCATTGAGCTTCCAATCAAAACACCCGGAACTGTTTGA    605

Qy   561  AAGTCTTGGAATTGCACACCAAAGGGTGTTGTCGTCTCTATGGCCACCTTGGTTACAGGTAA    620
          | |||||| ||||| |  |||| ||| |||| |||  |  || |||||| |||||||||| |
Db   606  GAGCCTTGGAATTGCACAACCAAAAGGGGTTCCACCTTTCTTATTGGCCACTGTGGGGTGCTGG    665

Qy   621  AACATTGTTGGCTAGGGGCTCATCATACTGACTGACTACATTCATTCAGGGTGTCTGG      680
          |  ||| ||||||| ||| |||||| ||| || ||||| |||||| ||||| ||||||
Db   666  GACATTGGTGGTTCAGAGTAATTGGTGAGGGCCTCCGGATGGGAACCTTCACTCAGGGTCTGG    725

Qy   681  TTCTGAGTTAGTTCAGAAAATACATTGAGAGAAGGTTCTAGAATGGGTCAGGGAAACTTTTTGT    740
          |||| || |||||| ||   ||| || || |||| | ||| |||| | |||| ||| |||||
Db   726  TTCTGAGTTGGTTCAGAGAGTATTGGGAGGGCCTCCGGATGGTTAGGGAGGTTAGGGAAACACTTT    785

Qy   741  TATGGCCAGGGGACATGCTCCATCAATTATCTTCATCGGATGAAAATTGACAGTAATTGGATC    800
          |  |    |||| ||| || |||| | || |  ||||| ||||| ||||| | |||| ||||
Db   786  TATGGCCAGGGAACACTCGCCACCATCAATATATTAGGGATGAAAATTGACTCTATCATATGCTGG    845

Qy   801  TGCTCGGATGGAATCTGGAAGTGGCAAGGGTGATTACAGCGTACTATATGCCTGGA    860
```

```
Db   846  TGCTAGAATGGAGCTGTCGGAACTCGGCAAACTCGACAACACGGTCACAGCGGTGAACGGAGTCACGGCCAC  905
          |||||  |||||||||   |||||| ||||| | |  |||||||||| | |   | | || | || |||| | |
Qy   861  ACTTCTCAACCAGTTGGAATGGACATTGAAAGCTTCA..AAATAAAGATCTCAAAAATAAGAGAGTTCAAGTTAAGGTTTTGATGGCCAC  920
          |  |||  |||||| ||| | |  ||| || || |||  ||   | || || |||||  || | | | ||||||||  ||||| ||| ||
Db   906  ACTTCTAAAACCAGCTTGATGGGTTTTTGAAGCATCAAAAACAAAGTTAAGGTTTTGATGGCCAAC  965

Qy   921  CAATCGGATTGATATCTCGGATCAAAGCCCTCCTTAGACCAGCACGAAGTAGACCGGAAAAT  980
          | ||| | ||| |    |||| |||||||   |  || ||| | |||||||| | |||||||
Db   966  GAACGAATAGACATTTTCGGATCAAGGCCCTTCTCGAGGCCGCCATAGACAGGAAGAT  1025

Qy   981  TGAATTCCAACCCCTAATGAAGAGTCTCGGCTGGATATTTTGAAAATCCATTCTAGAAG  1040
          |||||| |||| | |  |||| ||| ||||| || ||| | ||| || |||| ||| |
Db  1026  TGAATTTCAAAATCCTACGAGAATCCACGTTCGATATCTTGGAGGCCTTCGAACTCAAGAA  1085

Qy  1041  AATGAATTTAATGGGTGGGCATTGAATTGATTTGAAGAAGAGATTGCCGAGAGGATGAATGGCATC  1100
          || |  || ||||| |  || || ||| | |  || ||||||||   || || || || | ||||
Db  1086  AATGAACTTGATGGGTGGGCATTCGATCTCGAAAGAAGAGGATTCCGGAGAGGATGAATGGGGCCTC  1145

Qy  1101  TGGTGCTGAACTTAAGGCTGTTGCACTGAAGCGTGGAATGTTTGCTTTGAGGCGACGGAG  1160
          ||||||||||||| || ||||| |||||||| |||| || || ||  ||||| |  | ||
Db  1146  AGGAGCTGAGCTCAAGGCCCGCTCCACAGAGCCTGGAATGTTTTGCTTCGCTCGTGAGGAAG  1205

Qy  1161  GGTACACGTGACTCAGGAGGATTTTGAGGATGGGCCGTGGCGAAGGTGATGAAAAAGGAGAC  1220
          || | ||||| || ||||| |  | |||| ||| |  ||||||||| | || |||||| | |
Db  1206  GGTCGCACGGTTACCCAGGAGGACTTCGAGGATGGCCGCGATGAGATGGAGAGAAAGACAC  1265

Qy  1221  TGAAAAAAACATCTCATTGCCGGAAGTTGTGGGAG  1254
          ||| ||| |  | || ||   ||||| |||| |||
Db  1266  GGAGGAGAAACATGTCCCCTCCGCAAGCTCTGGGAG  1299
```

RESULT 2

```
BG645412
BG645412
LOCUS       BG645412       858 bp    mRNA    linear   EST 24-APR-2001
DEFINITION  EST507031 KV3 Medicago truncatula cDNA clone pKV3-46M11 5' end,
            mRNA sequence.
ACCESSION   BG645412
VERSION     BG645412.1  GI:13780524
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
            eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 858)
  AUTHORS   VandenBosch,K., Endre,G., Hur,J., Beremand,P., Town,C.D., Bowman,
            C.L., Craven,M.B., Cho,J. and Fraser,C.M.
  TITLE     ESTs from roots of Medicago truncatula 72 h after Rhizobium
            inoculation, 2001
  JOURNAL   Unpublished
COMMENT     Contact: VandenBosch K
            Department of Plant Biology
            University of Minnesota
            220 BioSci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
            Tel: 612 624 2755
            Fax: 612 625 1738
            Email: kvandenb@cbs.umn.edu
            M393819e TIGR sequence name: M'ECA78TK More information is
            available at: www.medicago.org
            Seq primer: Skmod (CTA gAA CTA gAT CC).
FEATURES             Location/Qualifiers
     source          1..858
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /cultivar="genotype A17"
                     /db_xref="taxon:3880"
                     /clone="pKV3-46M11"
                     /tissue_type="Seedling roots"
                     /dev_stage="3 days post-inoculation with Sinorhizobium
                     meliloti"
                     /lab_host="E. coli strain XLOlR"
```

RESULT 3

```
BASE COUNT        251 a    163 c    207 g    237 t
ORIGIN
Query Match     53.0%;  Score 664.6;  DB 10;  Length 858;
Best Local Similarity   88.8%;
Matches  731;  Conservative   0;  Mismatches  89;  Indels   3;  Gaps   1;

Qy        1 ATGGCTCTTGTTAGGAGTTGAACTGAACCATGCGGCGGCGTACCGGAGGGGGAATTGC 60
            ||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db       31 ATGGCTGCCGTCGGAGTAGAAGTGAAGCAT---GCCCGGAGGGGTGCACCGGAGGAGCAATTGC 87

Qy       61 TCCGCCAAGCCCACCAAGCAGCAGCGGAGGGCCTCCGCCACTACTAATTCTCCAACATCCAC 120
            ||||||||||||||| ||||||||||||| ||||||||||||||||||||||||||||||||
Db       88 TCTCGGAAACCAACCAAGAACAAGCGAGGGGTTCCGTCAATACTATACTCAGCCACATCCAT 147

Qy      121 GAGGCATCAGCTCCTTCTTCCGCCAAAAGACTCATACCTCAAGCGTCTCGAGGGCTCAGAGA 180
            |||||| |||||| |||| |||||||||| |||| |||| |||||||||||||||||||||||
Db      148 GAGCTTCAGCTCCTTCAACGTCAAAAGACTCATAATCTTAACCGTCGTCTGGACGCGCAACGT 207

Qy      181 AACGAGCTTCAATTCTAGGGTGAGGGAGGAATGCTGCCGAGAAGAATTACAGCTTCTGCAGGAACCC 240
            |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db      208 AACGAGCTCAATTCTAGGGTGAGGGATGACTTCGGGAGGAACTGCAGCTTCTTCTTCAGGAGGCCT 267

Qy      241 GGCTCTTATGTCGGTGAAGTTGTCAAAGTAATTGGCAGAACANAAAGTCCTTGTCAAGGTC 300
            | ||| |||||||||| ||||||||| ||| ||||| ||  | ||||||||||||||||||||
Db      268 GGTTCATATGTTGGTGAAGTCGTTAAAGTGATGGGCCAGAACAAAAGTCTTGGGTCAAGGTC 327

Qy      301 CACCCAGAAGGAAAAATATGTTCTTGTGACATTGACAAAAAATTTGCATTACAAAAGAATTACT 360
            ||||||| ||||||||||||| |||||||||||||||| |||||||||||| |||||||||||||
Db      328 CATCCAGAAGGGAAAATATGTTGTTGATAATTGGATTAAAAAAATATTGACATCACAACAGGATTACT 387

Qy      361 CCATCCACTAGAGACTTGCACTCCGCAACGACAGTTTATGTTCTTCACTTAGTTCTTCTGCCAAGT 420
            |||||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      388 CCATCTACTAGAGTTGCTCCTCGGCAACGACAGTTAGTGTCGTTCATTAGTTCTTCCCAGC 447

Qy      421 AAAGTTGATCCATTGGTCAATCTGATGANAAGTTGNGAAAGTTCCCGANTTCTACATATGAC 480
            |||||||  |||| ||||||||||| || |||||| |||||||||||| ||||||||||||||
Db      448 AAAGTTGATCCATTGGTTAATCTGATGAAAGTTGAGAAAGTTCCTGATTTCTACATATGAC 507

Qy      481 ATGATTGGTGGTTTAGACCACGACAAATTAAAGAAAATAAAAGAGGTCNTTAGACCAAC 540
            |||||||||||| ||||||||||| ||||||||||| ||||||||||| | |||||||||
Db      508 ATTGAGTCTCCAAT-TAACATCCTGAGTTGTTCTGAAAATTAGCTCAACCAAAG 567

Qy      541 AAACATCCTGAGTGTTCGANAGTTCNNANCCNAAGGGTGNGTCTGCTCTAT 600
            || ||||||||| ||||||| ||||| | | | ||||| | ||||| |||||
Db      568 AAACATCCTGAGTGTTCGAATTGCACAAACCCAAAGGGTGTGTCTGCTCTAT 627

Qy      601 GGGCCCACCTGGTACAGTGAAAAACATTGTTGGCTAGGGCGCTCATCATACTGAACTGT 660
            |||||||||||||||||||||| |||||||||||||||| |||||||||||||||||||
Db      628 GGGCCCACCTGGTACAGTGAAAAACAATTGTTGGCTAGGGCAGTAGGAGGTGCAGTAGCA 687

Qy      661 ACATTCATCAGGGTGTCTGGGTTCTGAGTCAGAAATACATTGGAGAAGAAGGTTCTAGA 720
            | | || |||||||||||| ||| |||| ||||| ||||||||||||||| |||||||||
Db      688 ACATTCATCAGGGTGTCTCGAGTTCAGAAATCATCAGGGTGTCCGGTTCTGAGTTGTTCAGAAATACAATC 747

Qy      721 AGGCTCAGGGAACTTTTTGTTATNGGCGAGGACACATGCTCCATCAATTATATCTCATGGAT 780
            |||||||| |||||||||||||| ||||| ||||||||||||||||||||||||||||||||
Db      748 ATGGTTAGGGAACTTTTTTGTTATGGCCAGGAGCATCCCCATCAATTATATCTCATGGAC 807

Qy      781 GAAATTGACAGTATTGGATCTCCTCGGATGAGGAATCTGGAAGTG 823
            |||||||||||||||||| |||||||||||||||||||||| ||||
Db      808 GAAATTGACAGTATTGGATCTCCTCGTATGGAAATCTGGAAGTG 850
```

```
AJ497641
LOCUS       AJ497641               723 bp    mRNA    linear
DEFINITION  AJ497641 MTFLOW Medicago truncatula cDNA clone mt--abc95112a08,
            mRNA sequence.
ACCESSION   AJ497641
VERSION     AJ497641.1  GI:22088084
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 723)
  AUTHORS   Firnhaber,C., Bartelsmeier,V., Becker,A., Meyer,F., Bartels,D.,
            Bekel,T., Linke,B., Puehler,A. and Kuester,H.
  TITLE     Determination of transcript sequences from developing flowers of
            Medicago truncatula genotype A17
  JOURNAL   Unpublished
COMMENT     Contact: Kuester H
            Lehrstuhl fuer Genetik
            Universitaet Bielefeld
            Postfach 100131 D-33501 Bielefeld, Germany.
FEATURES             Location/Qualifiers
     source          1..723
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /db_xref="taxon:3880"
                     /clone="mt--abc95112a08"
                     /tissue_type="flowers"
                     /dev_stage="different stages of development"
                     /clone_lib="MTFLOW"
                     /note="Vector: pGEM-T; Site_1: T-overhang; Site_2: XhoI;
                     genotype A17; cDNA was prepared from polyA+ enriched RNA
                     from flowers harvested at different stages of development.
                     The cDNA was directionally ligated into the pGEM-T vector
                     from Promega. Plasmids containing cDNA inserts were
                     propagated in E. coli DH5alpha cells."
BASE COUNT        216 a    118 c    188 g    201 t
ORIGIN
```

```
Query Match     50.4%;  Score 632;  DB 9;  Length 723;
Best Local Similarity   93.0%;
Matches  673;  Conservative   0;  Mismatches  50;  Indels   1;  Gaps   1;

Qy      466 GATTCTACATATNGACAATGCAGATTGGTGGTTTAGACCACAGACAAATTAAAGAAAATAAAAGAGGTC 525
            |||||||||||| |||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        1 GATTCTACATATGACACATGATTGGTGTTTAGACCACAACAAAATCAAAAGAAAATAAAAGAGGTC 60

Qy      526 ATTGAGCTACCAATCAAACATCCTGAGCTGTGTTGAAAGCTTGGAATTGCCACAACCAAAG 585
            ||||||||||||||||||| |||||||||| |||||||||||| |||||||||||||||||
Db       61 ATTGAGCTTCCAAT-TAACATCCTGAGTTGTTCTGAAAGTCTTGAAATAGCTCAACCAAAG 119

Qy      586 GGTGTCCTGCTCTATGGGGCCCACCTGGTACAGGGTAAAACATTGTTGGCTAGGGCGGCT 645
            ||||| |||||||||||| |||||||||||||||||| |||||||||||||||||| ||||
Db      120 GGTGTCTTGCTTTATGGGGCCCACCTGGTACAGGGTAAAACATTGTTGGCTAGGGCAGTAGCA 179

Qy      646 CATCATACTACGTTACATTCATCAGGGTGTCTGGTTCTGAGTTAGTTCAGAAATACATT 705
            ||||||||| |||||||||||||||||||||| ||||||||| |||||||||||||||||
Db      180 CATCATACTGATTGTTACATTCATCAGGGTGTCTCGGTTCTGAGTTGTTCAGAAATACATC 239

Qy      706 GGAGAAGGTTCTAGAAATGCTCAGGGAACTTTTTTGTTATAGGCGGGAACTGCTCCATCA 765
            |||||||| |||||| |||||||| |||||||||||||||| |||||||||||||||||||
Db      240 GGAGAAGGTTCTAGAAATGGTTAGGGAACTTTTTTGTTATGGCCAGGAGCATCCCCATCA 299

Qy      766 ATTATCTTCATGGATGAAAATTGACAGTATTGGATCTCGTCGGATGGAATCTGGAAGTGGC 825
            ||||| |||||| ||||| ||||||||||||| ||||| |||| |||||||||||||||||
Db      300 ATTATCTTCATGGACGAAATTGACAGTATTGGATCTCGTCGTATGGAAATCTGGAAGTGGC 359

Qy      826 AACGGTGATAGTGAGGTACACGGTACTATGCTCGGAACTTCTCAACCAGTTGGANTTGGATGGATT 885
            || | ||| |||||||| |||||| |||| |||||||||||| |||||||||| |||||| |||||
Db      360 AATGGTGATAGTGAAGTGCAGCGGTACTATGCTCGGAACTTCTCAACCAGTTGGATGGATGGGATC 419
```

```
Qy        886 GAAGCTTCAAATAAGATCAAGGTTTGATGGGCCACCAATCGGATTGATATCCTGGATCAA 945
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 GAAGCTTCCAATAAGATCAAGGTTTGATGGGCCACCAATCGGATTGATATCCTGGATCAA 479

Qy        946 GCCCTCCTTAGACCAGGAGGGATAGACGGGAAAAATTGAATTCCAACCCCTAATGAAGAG 1005
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 GCACTTCTTGAGACCTGGAGGGATTGACGGGAAAAATTGAATTCCAATCCTAATGAAGAG 539

Qy       1006 TCTCCGGCTGGATATTTTGAAAAATCCATTCTTAGAAGAATGAATTTAATTCCGGTGGCA 1065
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 TCTCGCCGGATATCTTGAAAAATCCATTCAAGAAGAATGAATTTAATGCGGTGGTATTGAT 599

Qy       1066 TTGAAGAAGAATTCCGCGGAAGAATGGGATGCAATCTCGGTTCGTCGTCGTCTGAACTTAA 1125
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 TTGAAGAAGAATTGTGCTGAGGAAGATGAATGAAGAGGGTTACACGTGCTGCAGCCTTGTGTGC 659

Qy       1126 ACTGAAGCTGGGAATGTTTGCTTTGAGGGAGCCGGAGGGGTTCACCGTGACTTCAGGAGGGATTT 1185
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 ACTGAAGCTGGAATGTTGCTTTGACGTGACACACGAGAGGATTT 719

Qy       1186 GAGA 1189
              ||||
Db        720 GAGA 723
```

RESULT 4
LOCUS   AK010505   1302 bp   mRNA   linear   HTC 05-DEC-2002
DEFINITION   Mus musculus ES cells cDNA, RIKEN full-length enriched library,
                clone:2410015M02 product:protease (prosome, macropain) 26S subunit,
                ATPase 5, full insert sequence.
ACCESSION   AK010505
VERSION   AK010505.1 GI:12846000
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
  PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
                Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
                prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
  PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
                Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
                Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
                Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
                Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
                Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
                Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE   RIKEN integrated sequence analysis (RISA) system-384-format
                sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
  PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
                Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
                Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
                Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
                Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
                Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
                Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
```

Second column:

```
                Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
                Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
                Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
                Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
                Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
                Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
                Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
                Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
                Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
                Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
                Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
                and Hayashizaki,Y.
  TITLE         Functional annotation of a full-length mouse cDNA collection
  JOURNAL       Nature 409 (6821), 685-690 (2001)
  MEDLINE       21085660
  PUBMED        11217851
REFERENCE       5
  AUTHORS       The FANTOM Consortium and the RIKEN Genome Exploration Research
                Group Phase I & II Team.
  TITLE         Analysis of the mouse transcriptome based on functional annotation
                of 60,770 full-length cDNAs
  JOURNAL       Nature 420, 563-573 (2002)
REFERENCE       6  (bases 1 to 1302)
  AUTHORS       Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
                Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
                Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
                Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
                Kaisukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
                Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
                Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
                Saito,H., Saito,R., Sakai,K., Sakai,Y., Sano,M., Sasaki,D.,
                Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
                Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
                Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
                Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE         Direct Submission
  JOURNAL       Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
                Physical and Chemical Research (RIKEN), Laboratory for Genome
                Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
                RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
                Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
                URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
                Fax:81-45-503-9216)
COMMENT         Please visit our web site (http://genome.gsc.riken.go.jp/) for
                further details.
                cDNA library was prepared and sequenced in Mouse Genome
                Encyclopedia Project of Genome Exploration Research Group in Riken
                Genomic Sciences Center and Genome Science Laboratory in RIKEN.
                Division of Experimental Animal Research in Riken contributed to
                prepare mouse tissues. First strand cDNA was primed with a primer
                [5'-GAGAGAGAAGGAAGATCCAAGAGCTCTTTTTTTTTTTTTTVN-3'], cDNA was
                prepared by using trehalose thermo-activated reverse transcriptase
                and subsequently enriched for full-length by cap-trapper. Second
                strand cDNA was prepared with the primer adapter of sequence [5'-
                GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3'], cDNA was cleaved
                with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
                Host: SOLR.
FEATURES             Location/Qualifiers
     source          1..1302
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:2410015M02"
                     /db_xref="MGI:1898859"
                     /db_xref="taxon:10090"
                     /clone="2410015M02"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /cell_type="ES cells"
     CDS             27..1247
                     /note="unnamed protein product; protease (prosome,
                     macropain) 26S subunit, ATPase 5 (MGP|MGI:105047)
                     putative"
                     /codon_start=1
```

```
                /protein_id="BAB26990.1"
                /db_xref="GI:12846001"
                /db_xref="MGI:105047"
                /translation="MALDGPEQMELEEGKAGSGLRQYYLSKIEELQLIVNDKSQNLRR
                LQAORNELNAKVRLLREEOLLQEOGSYGEVVRAMDKKVLVKVHPEGKEVVDVDKN
                IDINDVTPNCRVALRNDSVTLHKILPNKVDPLVSLMMVEKVPDSTYEMIGGLDKQIKE
                IKEVIELPVKHPELFEALGIAOPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSE
                LVQKFIGEGARMVRELFVMAREHAPSIIFMDEIDSIGSSRLEGGSGGDSEVQRTMLEL
                LNQLDGFEATNKIVIMATNRIDILDSALLRPGRIDRKIEFPPNEEARLDILKIHSR
                KMNLTRGINLRKIEELMPGASGAEVKGVCTEAGMYALRERGHVTQEDFEMAVAKVMQ
                KDSEKNMSIKKLWK"

     polyA_signal    1282..1287
                /note="putative"
     polyA_site      1302
                /note="putative"

BASE COUNT       347 a     285 c     388 g     282 t
ORIGIN
```

Query Match 50.3%; Score 630.2; DB 11; Length 1302;
Best Local Similarity 71.5%; Pred. No. 1.6e-156;
Matches 843; Conservative 0; Mismatches 333; Indels 3; Gaps 1;

```
Qy    76 AAGCAGGGACGGAGGGCCTCCGCCACTACTTCTCTCCAACATCCACGAGCATCAGTCCTT 135
            ||||||||||||||||| |||   ||| ||||||||||||||||||||||||||||||
Db    69 AAGGCAGGCAGTGGACTCCGTGCAAATATTATCTGTCTCAAGATTGAGGAACTCCAGTTGATT 128

Qy   136 CTTCGCCAAAAGACTCATTAACCTCAACCGTCTCGAGGCTCAGAGAAACGACCTCAATTCT 195
Db   129 GTAAATGATAAGAGCCAGAAATCTCCGTAGACTGCAGGCACAGAGGSATGAGCTGAATGCA 188

Qy   196 AGGGTGAGGAATCCTGCCCGAGAGAATTACAGCTTCTCGCAGGNACCCGGCTCCTTATGTCGGT 255
Db   189 AAAGTTCGCCTCGTTGCGGGAGGAGCTGCAGCTGTTGCACAGAAACAGGGCTCCTACGTTGGA 248

Qy   256 GAAGTTGTCAAAGTAATAGGGACAAGAGCCCTTGTCGTCAGGTCCACCCAGGAAGGAAAA 315
Db   249 GAAGTCGTGAGGGCCATGGATAAGAAAAAGTATTCGGTCAAGGTTCATCCTGAGGGCAAA 308

Qy   316 TATGTTGTTGACAATTGACAAAAATATTGACATTACAACAGATTACTCCATCCACTAGAGTT 375
Db   309 TTTGTTGTTGATCGGGRCAAGAAACATTGATAATCAACGAGTGCAGCGCCCAATTGTCGGGTC 368

Qy   376 GCACTCCCACCACAGTTATGTTCTTCACTTAGTTCTGCCAAGTAAAGTTGATCCATTG 435
Db   369 GCTCTAAGAAATGACAGCTACACTCGTCCATAAGATCTTTAACCTAACAAAGGTGGACACCCTTTG 428

Qy   436 GTCAAATCCTGATGAAAAGTTGAGAAAGTTCGCAGTTCCCATTCTTCACATATGCACATGATTTGGGTGGTTTA 495
Db   429 GTGTCACTAAATGATGGTTGGAGGGAAGGTGCCCAGACTCAACCTACGAGATGAITTGGCCGCGCCTG 488

Qy   496 GACCAGCAAATTAAAGAAAATAAAAGAGGTCATTGAGCTACCCAAATCAAACATCCTGAGCTG 555
Db   489 GACAAGCAGATCCAAGGAGAGTTAAAGAGTGTATGACGGCTGCCCCGTGAAGCACCCCGGAGCTC 548

Qy   556 TTTGAAAGTCTTGGAATTGCACAACCAAAGGGTGTCCCGTTCTATGGGGGCCCACCTGGTACA 615
Db   549 TTTGAAGCACTCGGGCATCGCAGCCAAAAGGGAGTCCTGCTTACGGACCCCACAGGCACT 608

Qy   616 GGTAAAACATTGTTGGCTAGGCGCAGTGGCTCATCATACGTGTCATTCATCAGGGTG 675
Db   609 GGGAAGACATTGTTGCGCCGGCCCTGTGGGCTCATCATACGAGACTGTACCTTTATTCGGTCGTC 668

Qy   676 TCTGGTGTTGTTCAGGAAATACATTGGAGAAGAGGTTCTTAGTAATGGGTCAGGGAACTT 735
Db   669 TCTGGCCTCTGAACTGGTTCAGAAATTCATCCGGGAGAGGGGCAAGAAATGGTTGAGGGAGCTG 728

Qy   736 TTTGTTATTGGGCCAGGGAACACTGCTCCATCAATTACTTCAGTGAGATGAAAATTGACAGTATT 795
Db   729 TTTGTTCATGGGCCCCGAGAACATGCTCCATCCATCCTTCATGGACCAGATTGACTTCATT 788

Qy   796 CGAATCTGCTCTGGATGAGGAATCTCGGAAGGTGGCAACGCGTTGATAGTAGTGAGGTACCAACGGTACTATG 855
Db   789 GGCATCCCTCACGGCCTGGAGGGGGGGCGGGCTCTGG---AGGGCCACAGTGAGGTACAGCGCGACGATG 845
```

```
Qy   856 CTGGAACTTCTCAAACCAGTTGGATTGGATTTGAAGCTTCAAATAACATCAAGGTTTTGATG 915
Db   846 CTGGAACTGCTCAATCAGCTGGATGGCCTTGAGGCCCAGGCCCACCAGGAGAATATCAAGGTTATCATG 905

Qy   916 GCCACCAATCGGATTGATATCCTGGATCAAGCCCTCCCTCCTTAGAACCACGAGCGGATAGACCGG 975
Db   906 GCTACCAATAGGATTGATAATCCTGGACTCTCCCCTCCTTCGTCCTCCGTCGGAGGATTCACAGA 965

Qy   976 AAAATTGAAATTTCCAACCCCTTAAATGAAGAAGAGTCTCCGGCTGGATATTTTGAAAATCCATTCT 1035
Db   966 AAAAATTGAATTCCCACCCCCTCCCAACGAGGAGGCCCCGGCCTGGACATTTTGAAAATCCACTCT 1025

Qy  1036 AGAAGAATGAAATTAATGCGTGGCATTGATTTGAAGAAGAATTGCCGAGAACGATGAATGGA 1095
Db  1026 CGGAAAAATGAACTTGACCAGGGGGGATCARACCTGAGGAAAAATTGCTGAGTGAGTTGCCGGGA 1085

Qy  1096 GCATCTGGTGTGTCTGAACTTAAGGCTGTTTGCACTGAAGCCTGGAATGTTTTTGCTTTGAGGGAG 1155
Db  1086 GCTTCAGGGCGTGAAGTGAAGGGGTGTGTGTGCACAGAAGCCTGGCCATGTACGCCCCTGCCGGAA 1145

Qy  1156 CGGAGGGGTTACAGGTGACTCACGAGNGGAGRTTTGAGGATGGCCGTGGGCCGAAAGGTGATGAAAAAG 1215
Db  1146 CGGGGGTGTGTCCACGTCACTCAGGAGCAGGACTTTGAGAGTGGCAGTTAGCTAAGGTCATGCAAAAG 1205

Qy  1216 GAGACTGAAAAAAACATGTCATTGCGGAAAGTTGTGGAAG 1254
Db  1206 GATAGTGAGAAAAAACATGTCCATCAAGAAAGTCATGAAG 1244
```

This clone has the following problem: frame shifted.

```
FEATURES             Location/Qualifiers
     source          1..1649
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:5537272"
                     /tissue_type="Embryo, stage 31/32, Xenopus"
                     /clone_lib="NICHD_XGC_Emb4"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
BASE COUNT     484 a    335 c    446 g    384 t
ORIGIN

Query Match  49.6%;  Score 622.6;  DB 11;  Length 1649;
Best Local Similarity  71.3%;
Matches  836;  Conservative  0;  Mismatches  334;  Indels  3;  Gaps  1;
```

```
Qy     82 GGGCAGGGGCCTCCGCCACTACTAATTCTCTCACATCCTCAACAGCTCTTCTTCGC 141
             || |  ||| |||  || ||| | ||| ||| || |  | |  ||||  || |||
Db    133 GGCATAGGCCTGCGGCAGTAATTACTCATCCACGAGCATCAGCTCAGCTTGTTGTTAAT 192

Qy    142 CAAAAAGACATCATAACCTCAACCGTCTCGAGCGTTCAGAGAAAACGACCTTCAATTCTTAGGGTG 201
            || |  | | ||| | || | | | |||| || || |||| ||| ||| ||  ||||   ||| ||
Db    193 GATAAGAGTCAGAATCTCAGGCGTTCTGCAAGCACAGAGGAAATGAACTGAACGGCTAAAGTG 252

Qy    202 AGGATGCTGCGCGGAAGAATTACAGCTTCTCTGCAGGGAACCCGGCTTCTTATGTCGGGTGAAGTT 261
            || ||||||  | ||||   |   | | || |||| |||||  | ||| | |  |  ||| |||||
Db    253 CGGGTGCTGCTGAGGGAGGAGGTGCAGCTGCCTTCAAGAGACGAGTACGTAGGGAAGTG 312

Qy    262 GTCAAAGTAAATGGGGCAAGAGAACAAAGTCCTTGTCAAGGTCCACCGAGAGGAAAAATATGTT 321
            |||| ||  || |  | |||| |||| ||| | ||  | |||| | |||| ||||||| | ||||
Db    313 GTCAGAGCTTATGCAACAAGAAGAAAGTAATGGGTTCAAGGTTCACCCCGAAGGGAAAATTCGTT 372

Qy    322 GTTGACATTGACAAAAATATTGACATTACAAAGATTACTCCCACTAGAGTTGCACTC 381
            || || | ||| | ||| ||||||| |  | |  || || |  || ||| |||| |
Db    373 GTCGATGTTGATAAGAATATTGACATCAATGAGTTCACAGCCAAACTGTCTGAGTTGCTCTG 432

Qy    382 CGGAACGACAGTTATGTTGTTCTTCACTTAGTTCTTCGCCAAGTAAAGTTGATCCATTGGTTCAAT 441
            ||||  |||||||| |  | | |   | |||||   ||||  | | |  |   ||  || ||||| |
Db    433 CGGAATGACAGTTACACCGCTGCAATAAGATCTCCCCCAACAACAAGGTGGGACCCCTTAGTTCTCT 492

Qy    442 CTGATGAAAAGTTGGAGAAAGTTCCCGATTCTACATATGCATCATGATTGGTGGTTTAGACACAG 501
            |||||||  |||| || ||||||||  || | ||  |  | |||||| |||| || |  ||| ||
Db    493 CTGATGATGGTGAGAAAGTTCCAGACTCCACGTACAGACAATTGTGGTGTTTGGGATTAA 552

Qy    502 CAAATTAAGAGAATAAAAGAGAGGTCATTGAGCCTACCAATCAAACATCCTGAGCTGTTTGAA 561
            ||||| |||| || ||||| || |||||||| ||| || | |||||| || ||||| ||||||
Db    553 CAGATAAAAGGAAAATCAAGGAGGTCATTGAACCTCGCCTGTTGAAAACACCCAGAGCTCTTTGAA 612

Qy    562 AGTCTTGGAATTGGACAACCAAAAGGGTGTCCTCCTTCTTATAGGGCCACCTGGTTACAGGTAAA 621
             | ||||| | |  ||| ||||||||| | || | | |  || ||| ||||||| |||| | |||
Db    613 GCCTTGGGTTGATCCGCGGCCCCAAAGGGGCGTGTTGTGTACGGACCTCCTGGCACGGGAAAG 672

Qy    622 ACATTGTTGGCTAGGGCAGTGGGCTCATCATACTGACTGTACATTCATCAGGGTGTCTGGT 681
            ||| ||| ||| | || ||||| | || ||| || || |  |||| | | ||||||||||
Db    673 ACAACTTTGGCTCGTGCTGTGGCCCATACAGCTACAACATTTATACGAGTGTCTGGA 732

Qy    682 TCTGAGTTAGTTCAGAAATACACATTGGGAGAAGTTTCTAGAAATGGTCAGGGAGACTTTTGTT 741
            || ||| | || || |||| | || | ||| ||||| |  |   |||| ||  |||| |||||||
Db    733 TCCGAAACTGGTCGAGAAGTTCATCGGGGAAGAGTTGCGTGGTAAAGGTGCTGAACTGTTTGTT 792

Qy    742 ATGGCCAGGAACATGCTCCATCAAATTATCTTCATGGATGAAATTGACAGTATTGGATCT 801
            ||||| ||| ||||||||| ||| |||||  | | ||| |||||||||  | ||||| | |
Db    793 ATGGCTCGGAGAACATGCTCCCATCATTTATGTCAGGAAATTGATTCGAATTGGGTCG 852

Qy    802 GCTCGGATGGAAGTTGGCAACGGTGATAGTGAGGTGCAACGGTACGGTACTATGCTGGAA 861
            || ||| | ||||  ||   ||| |  |||||||| |||||| || | ||| ||||| ||
Db    853 TCGGGGCCTGAGGGGGGTCGGG---AGGAGACAGTGAGGTCCAGCGCACAATGCTTGAG 909

Qy    862 CTTCTCAACCAGTTGGATGGGATTGGAAGCTTCAAAATAAGATCAAGTTTGATGGGCCACC 921
            || || |||||| |||||||||||  | ||| ||||| || |||| | | |  ||||| ||
Db    910 CTGCTGAACCAACTGGATGGGATTTTGAGGCAACAAAAGAACATAAAGGTAATCATGGCAACT 969

Qy    922 AATCCGGATTGATATACCTGGATCAAGCCCTTCCTTAGACCAGGACCGGATAGACCGGAAAATT 981
            ||  |||||||| || || || ||| ||| |  ||| || | ||||| ||  |||||||| | |
Db    970 AACAGGAATTGAATTATATTGGAGACTCTGCCTTTTGCGTCGGGCCCAGGGTCGCATTGACCGGAAGATA 1029

Qy    982 GAAATTCCAACCCCTTAATGAAGAGTCTTCGGCTCGGCTGATATTTGAAAATCCATTCTAGAGA 1041
            ||||| || ||| ||||||||||| |    |||| | | |||||||||||   |||| | | ||
Db   1030 GAATTCCCACCTCTTAATGAAGAGGCCCGTCTGACACATTTGAAAGGTAAAA 1089

Qy   1042 ATGAATTAATGCGTGGCATTGATTTGAAGAAGATTGCCGAGAAGATGAATGAACTCT 1101
            |||||| || | |||||  | | |||| | ||||||| |||| ||||| || |  |
Db   1090 ATGAACTTGACGCGTGGGGATCAACCTGCCGCAAAATAGCAGAACTGAATGCCTGGAGGTCG 1149

Qy   1102 GGTGCTGAACTTAAGGCTGTTTGCACTGAAGCTGGAAATGTTTGCTTTGAGGGAGGCGGAGG 1161
            |  ||||||  ||||| |||||||| || |||| || | | |   | | ||| || || ||
Db   1150 GGGGCGTGAAGTAAAGGGGGGTGTTGCACAGAAGCGGGAAAGTATGGCCCTCCGTGAGGAGGCCG 1209

Qy   1162 GTACACGTGACTCAGGAGGATTTTGAGATGGCCGTGGCCAAGGTGATTGAAAAAGGAGACT 1221
            || ||| | ||| |||||||||||||| ||||| |||||||| | |  | || | ||| ||
Db   1210 GTGCACGGTCACAGAAGAGGACTTTGAAATGGCCTGGGCCAAAGGTCATGCAGAGAAGAGACACT 1269

Qy   1222 GAAAAAACAATGTCATTGGGAAGTTGTGGGAG 1254
            || |  ||| ||||| || ||||| ||||| |
Db   1270 GAGAGAACACATGTCCATTAAGAAGCTGTGGAAG 1302
```

RESULT 6
CA784406
LOCUS   CA784406   742 bp   mRNA   linear   EST 04-DEC-2002
DEFINITION sat98h03.y1 Gm-c1062 Glycine max cDNA clone SOYBEAN CLONE ID:
            Gm-c1062-6893 5' similar to TR:Q9SCD6 Q9SCD6 26S PROTEASOME SUBUNIT
            8. ;, mRNA sequence.
ACCESSION   CA784406
VERSION   CA784406.1 GI:26047693
KEYWORDS   EST.
SOURCE   Glycine max (soybean)
  ORGANISM Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
            eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1 (bases 1 to 742)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Mattin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,
            Y., Person,D., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,B., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE   Public Soybean EST Project
  JOURNAL   Unpublished
COMMENT   Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, Al 35801 For further information
            call: (800)-533-4363 or contact: ccu@resgen.com web site:
            www.resgen.com
            Seq primer: -40RP from Gibco
            High quality sequence stop: 438.
FEATURES   Location/Qualifiers
     source   1..742
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /db_xref="taxon:3847"
                     /clone="SOYBEAN CLONE ID: Gm-c1062-6893"
                     /tissue_type="stem tissue of greenhouse grown plants"
                     /dev_stage="1 month old"
                     /lab_host="DH10B"
                     /clone_lib="Gm-c1062"

RESULT 7

LOCUS   BI968213/c
DEFINITION   BI968213
             GM830004B12F11 Gm-r1083 Glycine max cDNA clone Gm-r1083-1437 3',
             mRNA sequence.
ACCESSION   BI968213
VERSION   BI968213.1 GI:16342618
KEYWORDS   EST.
SOURCE   Glycine max (soybean)

                      794 bp   mRNA   linear   EST 23-OCT-2001

Query Match   49.0%; Score 614; DB 14; Length 742;
Best Local Similarity   96.2%; Pred. No. 2.5e-152;
Matches 629; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

```
Qy          1 ATGGCTCTTCTTGTAGGAGTTGAACTGAAGCATCGGCGGCGGAGGGCGGAGGGCGGAATTGC 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         89 ATGGCTCTTGGTCGGAGTTGAAGCATCGCGGCGGAGGGCGCACCGAGGAGGAATTGC 148

Qy         61 TCCGCCAAGCCCAACCAAGCGAGGGCGGAGGGCCTCCGCCCACTACTTATTCTCTCAAACATCCAC 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        149 TCCGCCAAGCCCACCAAGCAGGGAGGGCCTCCGCCCACTACTATTCTCACAACATCCAC 208

Qy        121 GAGCATCAGCTCCTTCTTCCTCGCCAAAAGACTCATAACCTCAACCGTCTCGAGGCTCAGAGA 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        209 GAGCATCAGCCTCCTTCTTCCGTCAAAAGACGCATAACCTCAACCGTCTTGAGGCTCAGAGA 268

Qy        181 AACGACCTCAATTCTAGGGTGAGGATGCGTCCCGGACAGAAATTACAGCGTTTTGCAGGAACCC 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        269 AACGAGCTCAATTCCAGGGTGAGGATGCCTGCCGGAAGAATTACAGCTTTTGCAGGAACCA 328

Qy        241 GGCTCTTTATGTCGGTGAAGTGTCTCAAAGTAATGGGCAAGAACAAAAGTCCTTGTGTCAAGGTC 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        329 GGCTCTTTATGTTGGTGAAGTTGTCAAAGTCATGGGCAAGAACAAAGTCCTTGTCAAGGTC 388

Qy        301 CACCCAGAAGGAAAATATGTTGTTGACAAAATATTGACATTGACATTACAGAAGAATTACT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        389 CACCCAGAAGGAAAATATGTCGTTGATATTGACAAAAGTATTGACATTACAAAAGATTACT 448

Qy        361 CCATCCACTAGAGTTGCACTCCGCAACGACCAGTCTATGTTTCTTCACTTAGTTCTGCCCAAGT 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        449 CCATCCACTAGAGTTGCACTTGGCACCAGTTATGTTGTTCTTCACTTAGTTCTGCCAAGT 508

Qy        421 AAAGTTGATCCATTGGTGTCAATCTGATGAGAAGTTCCGGATTTTACTATATGAC 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        509 AAAGTTGATCCATTGGTCAATCTGATGAACAGTTGAGAGGTTCCGGATTCTACTTACGGAC 568

Qy        481 ATGATTGGTGGTTTAGACCAGCAGAAATTAAAGAGAATAAAAGAGGTCATCTGAGACCAATC 540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        569 ATGATTGGTTGGTTTAGACACAGCAGAAATTAAAGAGGGTCATTGAGGTTACCAATC 628

Qy        541 AAACATCCTGTAGGTAAAAAACATTGTTGTTGGCAGCACAACCAAAGGGGTGTCCTTGTTCTAT 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        629 AAACATCCTGAGCTGTTTGAAAGTCTTGGAATTGCACAACCAAGGGGTGTCTGCTTCTTAT 688

Qy        601 GGGCCACCTGGTTACAGGAGAAAAACATTGTTGGCTAGGGCCAGTGGGCTCATCATACT 654
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        689 GGGCCCACCTGGTTACAGGAGAAAAACATTTGTTGGGCTAGGGCCAGTGGGCTCATCATACT 742
```

BASE COUNT   226 a   156 c   185 g   175 t
ORIGIN

ORGANISM Glycine max
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
          ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
          Glycine.
REFERENCE   1 (bases 1 to 794)
AUTHORS   Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
            Erpelding,J., Raph,C., Shoop,E., Pardinas,J., Liu,B. and Lewin,H.
TITLE   A Functional Genomics Program for Soybean (NSF 9872565)
JOURNAL   Unpublished
COMMENT   Other ESTs: AI855730 corresponding to Gm-c1013-1501 (5')
            Contact: Vodkin,L.O., PI, A Functional Genomics Program for
            Soybean (NSF 9872565)
            Lewin, H. A., Director, Keck Center for Comparative and Functional
            Genomics
            University of Illinois
            Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
            Tel: (217) 244-6147
            Fax: (217) 333-4582
            Email: l-vodkin@uiuc.edu
            This clone is available through: Incyte Genomics, 4633 World
            Parkway Circle St. Louis, Missouri 63134. Phone (800) 430-0030 or
            (314) 427-3222 FAX: (314) 427-3324. Web site:
            http://www.incyte.com/reagents/catalog.jsp?page=clones/collaboratio
            n/index
            Seq primer: 5'TTTTTTTTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
FEATURES   Location/Qualifiers
     source   1..794
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /db_xref="taxon:3847"
                     /clone="Gm-r1083-1437"
                     /clone_lib="Gm-r1083"
                     /note="The library Gm-r1083 is a sequence-driven, reracked
                     set of 4,992 clones selected from cDNA libraries from
                     various tissues and stages of development of soybean. It
                     represents 1117 sequences from the progenitor library
                     Gm-c1009 (from mature roots of 2 month old greenhouse
                     grown 'Williams' soybean plants); 820 sequences from the
                     progenitor library Gm-c1013 (from 2 to 3 week old whole
                     plants of Williams); and 3055 sequences from library
                     Gm-c1028 (from 'Supernod' plants whose seedlings were
                     inoculated with Bradyrhizobium japonicum, courtesy of Dr.
                     Gary Stacey). The 5' ESTs of the source clones from the
                     different progenitor libraries was used to select
                     singletons, or a representative of each contig, which were
                     reracked to form library Gm-r1083. The cDNA clones of the
                     reracked Gm-r1083 library were then sequenced at the 3'
                     end. The contig analysis to select unique genes was
                     performed by the laboratory of Ernest Retzel, Center for
                     Computational Genomics and Bioinformatics, University of
                     Minnesota, http://web.ahc.umn.edu/biodata/nsfsoy/.
                     Reracking was performed by Incyte Genomics, St. Louis,
                     http://www.incyte.com, and 3' sequencing by the Keck
                     Center for Comparative and Functional Genomics, University
                     of Illinois, http://www.lie.uiuc.edu/biotech/keck.html.
                     Note: The corresponding 5' EST from each clone in the
                     Gm-r1083 library is listed in the 'OTHER EST' field. The
                     detailed information on the source library for each clone
                     can also be obtained by referring to the Incyte Genomics
                     clone ID of the original cDNA library that is also listed
                     under 'OTHER EST'."

BASE COUNT   224 a   211 c   129 g   214 t   16 others
ORIGIN

Query Match   48.5%; Score 608.4; DB 12; Length 794;
Best Local Similarity   97.3%; Pred. No. 8.1e-151;
Matches 609; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
Qy        629 TGGCTAGGGCCAGTGGGCTCATCATACTGACTGTACATTACATTCAGGGTGTGTCTGGTTCTCGAGT 688
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        792 TGNNNAGNNNAGTCGCTTCATCATACTGACTGTACCATTCAGTGACTATCATCAGGGTGTCTGNGTTCTCNGNTNNNGAGT 733
```

```
RESULT 8

LOCUS       CB979660      803 bp    mRNA    linear    EST 01-MAY-2003
DEFINITION  CAB70001_IIIaF_E11 Cabernet Sauvignon Berry Post-Veraison - CAB7
            Vitis vinifera cDNA clone CAB70001_IIIaF_E11 5', mRNA sequence.
ACCESSION   CB979660
VERSION     CB979660.1  GI:30302866
KEYWORDS    EST.
SOURCE      Vitis vinifera
  ORGANISM  Vitis vinifera
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; Vitaceae; Vitis.
REFERENCE   1  (bases 1 to 803)
  AUTHORS   Goes da Silva,F., Iandolino,A., Lim,H., Baek,J., Jones,K. and Cook
            ,D.
  TITLE     Expressed sequence tags from cabernet sauvignon berries at various
            developmental stages
  JOURNAL   Unpublished
COMMENT     Contact: Douglas Cook, PhD
            CAES Genome Facility
            UC Davis, Plant Pathology
            One Shields Ave, Davis, CA 95616, USA
            Tel: 530 754 6561
            Fax: 530 754 6617
            Email: drcook@ucdavis.edu
            Seq primer: ACGGTACCGGACATATGCC.
FEATURES             Location/Qualifiers
     source          1..803
                     /organism="Vitis vinifera"
                     /mol_type="mRNA"
                     /cultivar="Cabernet Sauvignon"
                     /db_xref="taxon:29760"
                     /clone="CAB70001_IIIaF_E11"
                     /clone_lib="CAB70001_IIIaF_E11"
                     /dev_stage="Post-Veraison, 18-19 brix"
                     /sex="Hermaphrodite"
                     /lab_host="DH5alpha"
                     /note="Organ: Berry; Vector: pDNR; Site_1: SfiI; Site_2:
                     SfiI; CAB7 is a cDNA library of Cabernet Sauvignon Clone 8
                     berries. Samples were collected post-veraison from
                     field-grown vines during stage III of berry growth at 89
                     days after full bloom. Berries soluble solid concentration
                     ranged between 18-19 brix. Sampled vines were located at
                     the University of California, Davis. Experimental
                     Vineyard. cDNAs were made by oligo-dT priming and directi
                     onally cloned. 5' and 3' adaptors were used in cloning as
                     follows: 5' AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3', and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Library was
                     constructed using the Clontech Creator SMART kit and
                     size-selected to contain the 0.5-3 kb size fraction."
BASE COUNT      238 a    147 c    194 g    224 t
ORIGIN
```

```
Query Match
Best Local Similarity   84.9%;   Pred. No. 8.1e-151;
Matches 681;  Conservative  0;   Mismatches 121;   Indels  0;   Gaps  0;
Score 608.4;  DB 14;   Length 803;
```

Left alignment:

```
Qy  689 TAGTTCAGAAATACATTGGAGAAGGTTCTPAGAATGGTCAGGGAACTTTTTGTTGTTATGGCCA  748
        |||||||||||||||||||||||||||    |||||||||||||||||||||    ||||||||
Db  732 TAGTTCAGAAATACATTGGAGAAGGTTNNNNAGAATGGTCAGGGAACTNNNNGTTATGGCCA  673

Qy  749 GGGAACATCCTCCATCATCAATTATTTCATGTCCGGATCAAAATTGACTAGTTGGATCTGTCTGCTCGGA  808
        |||||||| ||||||  ||||||| ||||||||||||||||||||||||||||||||||||||||
Db  672 GGGAACATGCTCCATCAATTATCTTCATGTCCGGATCAAAATTGACTAGTTGGATCTGCTCGGA  613

Qy  809 TGGGAATCTGGAAGTCGGAACGGTGATAGTGAGGTACCACGGTACTATGTGGGAACTTCTCA  868
        ||| |||||||||||  |||||||||||||||||||||| ||||||||| |||||||||||
Db  612 TGGAATCTGGAAGTGGCAACGGTGATAGTGAGGTACACGGTACTATGCTGGAACTTCTCA  553

Qy  869 ACCAGTTGGATGGATTTGAAGCTTCAAAATAAGAATCAAGGTTTTGATGGCCACCAATCGGA  928
        |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db  552 ACCAGTTGGATGGATTTGAAGCTTCAAAATAAGATCAAGGTTTTGATGGCCACCAATCGGA  493

Qy  929 TTGATATCCTGGATCAAGCCCCTTAGCACCAGGACCGGAAAAATTGAATTTC  988
        |||||||||||||||||||||||| | |||||||||||||||||||||||||
Db  492 TTGATATCCTGGATCAAGCCCCTTCAACCAGGACCGGAAAAATTGAATTTC  433

Qy  989 CAACCCTTAATGAAGAGTCTCGGCGTTGGATATTTGAAAAATCCATTCTTAGAGAATGAATT  1048
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  432 CAACCCTTAATGAAGAGTCTCGGCGTTGGATATTTGAAAAATCCATTCTTAGAGAATGAATT  373

Qy 1049 TAATGCGTGGCATTGATTTGAAGAAGATTGCCCAGAAGATGAATGGGCATCTTGGTGCTG  1108
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  372 TAATGCGTGGCATTGATTTGAAGAAGATTGCCCAGAAGATGAATGGGCATCTTGGTGCTG  313

Qy 1109 AACTTAAGGCTGTGTTTTGCACTGAACGTGAACCTGGAAATGTTTGCTTTGAGGGAGGCGGGAGGGTACACG  1168
        |||||||||||| |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db  312 AACTTAAGGCTGTTTTGCACTGAACCTGAACCTGGAAATGTTTGCTTTGAGGGAGGCGGGAGGGTACACG  253

Qy 1169 TGACTCAGGAGGGGATTTTGAGATGGCCCGTGGCGAAGGTGATGAAAAAGGAGGACTGAAAAA  1228
        ||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  252 TGACTCAGGAGGATTTTGAGATGGCCCGTGGCGAAGGTGATGAAAAAGGAGGACTGAAAAA  193

Qy 1229 ACAGTCATTGCCGGAAGTTGTGGAAG  1254
        ||||||||||||||||||||||||||
Db  192 ACAGTCATTGCCGGAAGTTGTGGAAG  167
```

Right alignment:

```
Qy  345 CATTACAAAGAGATTACTCCATCCACTCAGAGTAGAAAACATTGTTGGCTAGGGCACGTGCC  404
        ||| || |||| |||| || | ||||||| || ||||| ||||| ||||||||| | |||||
Db    1 CATCACTAAGATCACACCTTCAACCCGAGTTGCTCTTCGCAATGAGATCTATGTGCTTCA   60

Qy  405 CTTAGTTCTGCCAAGTAAAGTTGATCCATTGGCAAGTCAATTGATGAAAGTTGAGAAAGTTCC  464
        ||| |||||||| |||||||  | |||| || |  | || ||||||| |||||||| ||||||
Db   61 CTTGATATTGCCCAGTAAAGCTAGATCCCACTCGGTAAACTTTATGGAAGTTGAAAAGTTCC  120

Qy  465 CGATTCTACATATGACATGATTGGTGGTTTAGACCAGCAGAATTAAAGAAAATAAAAGAGGGT  524
        ||||||||||||||| |||||||||||| |     |||||||| ||||| || ||| |||||
Db  121 GGATTCTACATATGATATGATTGGTGGTCTCTTGATCAGCAGAAATCAAAGAGGATAAAGGAGGT  180

Qy  525 CATTGAGCTACCAACTACAAACATCCTGAGCTGTGTTTGAAAGTCTTGGCTATAGACCTCAACCAAA  584
        |||||||||  || | ||  |||||||||  | | |||||||||||||||||||||||||||||||
Db  181 TATTGAGCTTCCAATCAAGCCATCCTGAATTATTTGAAAAGTCTTGGCTATAGACCTCAACCAAA  240

Qy  585 GGGGTCGTCCCGCCTCTATGGGCCCACCTGGGTACAGGGTAAAACATTGTTGGCTAGGGCAGTGCC  644
        |||| || |||||| |||||| || |||||||||||||| || ||||| ||||| |||||||||||
Db  241 GGGGAGTCCCTGCTTTATGGGTCCTCCTGGGTACAGGGAAGACACTTCTTGCAAGGGCAGTGCC  300

Qy  645 TCATCAATACTGACTGTACATTCATCAGGGGTGTCTGGGTTCTGAGTTAGTTCAGAAATACAT  704
        ||||| | |||||||||||  | ||  |||| | |  | |||||||||||||||||||   ||
Db  301 TCATCACACTGACTGTACCTTTATAAGGGTTCTCGGCTCGAGTTAGTTCAGAAGTTAT  360

Qy  705 TGGAGAAGGTTCTAGAAATGGTCAGGGAACTTTTTGTTGTTATGGCCAGGGAACATGCTCCATC  764
        ||| ||||||||| ||||||| |||| |||| ||||||| |   |||| ||||||||||||||||
Db  361 TGGTGAAGGTTCTCGAAATGGTTAGAGAACTGTTTTGTTAGTCAAGGGAACATGCTCCATC  420

Qy  765 AATTATCTTCATGGATGATGAAAATTGACAGTGATCGTGCCTCCGAGATGGAGGAAATCTGGAAGTGG  824
        ||| ||||||||||||||||||||||||||| || |  | ||||||||||||| ||||||||||||
Db  421 AATCATCTTCATGGATGATGAAAATTGACAGTAATGGAAGGGAGGACCATGCTCCATC  480

Qy  825 CAACGGTGATAGTGAGGTACAGCGTTACTATGTCGGAACTTCTCAACCAGGTTGGATGGATT  884
        ||| |||||||||||| || |||| |||||||| |||| ||||||||  || ||||||||||
Db  481 AAATGGTGATAGTGAAGTGCAGCGGAACTATGTGGAGCTTCTCAATCAACTGGATGGATT  540

Qy  885 TGAAGCTTCAAAATAAGATCAAGGTTTTGATGGCCACCAATCGGATTGATATCCTGGATCA  944
        ||||||  ||||| ||||| ||||||||||||||  || |||||| |||||| | ||||||
Db  541 TGAAGCCTCAAACAAGATTTTGATGGGCTACAAAATCGCATTGATATTTGGATCA  600

Qy  945 AGCCCTCCTTAGACCCAGGACGACGGAAAAATTGAAATTCCAACCCCTAATGAAGA  1004
        ||||||||||||| || |||||  | || |||||||| |||||||| |||| |||||
Db  601 AGCCCTCCTTAGGCCTCGGACGAAATTGACAGGAGAGATTGAGTTCCCAAAACCCAACGAGA  660

Qy 1005 GTCTCGGGCTGGATATTTGAAAAATCCACTCTTAGAAGAATGAAATTTAATGCGTGGCATTGA  1064
        |||| || ||| ||||||||||||||||||||  |||| |||||| | ||||||| || ||||
Db  661 GTCCCGCTTGAATATTTTGAAAAATCCACTCAAGAAAAAATGAACTTAATGCGCGGGGATTGA  720
```

```
RESULT 9
BE942284
LOCUS       BE942284      685 bp    mRNA      linear   EST 03-OCT-2000
DEFINITION  EST421863 MGHG Medicago truncatula cDNA clone pMGHG-8E11, mRNA
            sequence.
ACCESSION   BE942284
VERSION     BE942284.1  GI:10520043
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 685)
  AUTHORS   Cote,F., Ojanen-Reuths,T., Hahn,M.G., VandenBosch,K., Hur,J.,
            Beremand,P., Endre,G., Town,C.D., Bowman,C.L., Craven,M.B. and Cho
            ,J., Fraser,C.M.
  TITLE     ESTs from seedling roots of Medicago truncatula after treatment
            with beta glucan elicitor preparation from Phytophthora sojae
  JOURNAL   Unpublished
COMMENT     Contact: Michael G. Hahn
            Complex Carbohydrate Research Center
            University of Georgia
            220 Riverbend Road, Athens, GA 30602-4712, USA
            Tel: 706-542-4457
            Fax: 706-542-4412
            Email: hahn@cccrc.uga.edu
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
            More information is available at: http://chrysie.tamu.edu/medicago
FEATURES             Location/Qualifiers
     source          1..685
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /cultivar="A17"
                     /db_xref="taxon:3880"
                     /clone="pMGHG-8E11"
                     /tissue_type="Roots from four day old seedlings"
                     /dev_stage="2 days after treatment with beta glucan
                     elicitor preparation from Phytophthora sojae"
                     /lab_host="E. coli strain XLOLR"
                     /clone_lib="MGHG"
                     /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
                     was directionally ligated into the Unizap XR vector from
                     Stratagene and packaged using Gigapack III Gold packaging
                     extracts. Plasmids containing cDNA inserts were excised
                     from the recombinant lambda-Zap phage using Ex-assist
                     helper phage and propagated in XLOLR cells."
BASE COUNT     205 a    115 c    177 g    188 t
ORIGIN

Query Match       48.2%;  Score 605;  DB 10;  Length 685;
Best Local Similarity   92.7%;  Pred. No. 6.1e-150;
Matches   635;  Conservative   0;  Mismatches  50;  Indels   0;  Gaps   0;

Qy   490 GGTTTAGACCAGCAAATTAAAGAAATAAAAGAGGTCATTGAGCTACCAATCAAACATCCT 549
         ||||||||||||||||||||||||||||| ||||||||| ||||||||||||||||||||
Db     1 GGTTTAGACCAACAAATCAAAGAAATAAAACAGGTCATTGAGCTTCCAATTAAACATCCT 60

Qy   550 GAGCTGTTTGAAAGTCTTGGAATTGCACACCAAAAGGGGTGTCTTCCTGCTTCATTGGGGCCACCT 609
         |||| ||||| || |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    60 ...
```

```
Qy  1065 TTTGAAGAAGGATTCCCGAGAAGAGATGAATGAGAGCATCTCGGTGCTGTGTGTGAACTTAAGGCTGTTTG 1124
         ||||||||||| ||| |||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   721 TCTGAAGAAGGATTAGCAGAGAAGAGAAGAGRAATGGGTGCATCTCGGTGCAGAGGCTTAAGGCGGGTATC 780

Qy  1125 CACTGAAGTGGAATGTTTGCT 1146
         || |||||||||||||||||||
Db   781 CACAGAAGCTGGAATGTTTGCT 802
```

```
RESULT 10
AK012294
LOCUS       AK012294     1313 bp    mRNA      linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 11 days embryo whole body cDNA, RIKEN full-length
            enriched library, clone:2700027102 product:protease (prosome,
            macropain) 26S subunit, ATPase 5, full insert sequence.
ACCESSION   AK012294
VERSION     AK012294.1  GI:12849948
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
```

```
Qy    61 GAGTTGTTCGGAAAGTCTTGGGAATAGCTCAACCAAAAGGGGTGTCTTGCTTTATGGGGCCACCT 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 GAGTTGTTCGGAAAGTCTTGGGAATAGCTCAACCAAAAGGGGTGTCTTGCTTTATGGGGCCACCT 120

Qy   610 GGTACAGGTAAAAACATTGTTGGCTGACGGCCGTGGCTCATCATCATCAAGTTGCATTCATC 669
         ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GGTACAGGTAAAAACATTGTTGGCTAGGGCCGTGGCTCATCATCGACTGACTGACATTCATC 669

Qy   121 GGTCACAGGAAAAACATTGTTGGCTGAGGGCCGTAGCACATCATCATGATTGTACATTCATC 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GGTCACAGGAAAAACATTGTTGGCTGAGGGCCGTAGCACATCATCATGATTGTACATTCATC 180

Qy   670 AGGGGTGTCTGGTTCTGAGTTAGTTGTTCAGAAAATACATCCAGAAGAAGGTTCTAGAATGGTCAGG 729
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 AGGGGTGTCTGGTTCTGAGTTAGTTGTTCAGAAAATACATCCAGAAGAAGGTTCTAGAATGGTCAGG 729

Qy   181 AGGGGTGTCTGGTTCTGAACTTGTTCAGAAATACATCCAGAGAAGAAGGTTCTAGAATGGTTAGG 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AGGGGTGTCTGGTTCTGAACTTGTTCAGAAATACATCCAGAGAAGAAGGTTCTAGAATGGTTAGG 240

Qy   730 GAACTTTTGTTATGGCCACAGGGAACATGCTCCATCAATTATCTTCATCGATGAAATTGAC 789
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GAACTTTTGTTATGGCCACAGGGAACATGCTCCATCAATTATCTTCATCGATGAAATTGAC 780

Qy   241 GAACTTTTGTTATGGCCAGGGAGGCAGGGCCCCATGCAATTATCTTCATGGACGAAATTGAC 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GAACTTTTGTTATGGCCAGGGAGGCAGGGCCCCATGCAATTATCTTCATGGACGAAATTGAC 300

Qy   790 AGTATTGGATCTGCTCCGGATGGAGAATCTGAAAGTGGCAACGGGTGAGTTGATTAGTGAGGTTACAGGGT 849
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 AGTATTGGATCTGCTCCGGATGGAGAATCTGAAAGTGGCAACGGGTGAGTTGATTAGTGAGGTTACAGGGT 849

Qy   301 AGTATTGGATCTGCTCGTGGATGGAGAATCTGAAAGTGGCAATGGTGATAGTGAAGTGCAGCGT 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AGTATTGGATCTGCTCGTGGATGGAGAATCTGAAAGTGGCAATGGTGATAGTGAAGTGCAGCGT 360

Qy   850 ACTTATGCTGGAACTTCTCAACCAGTTGGATTGGAGCTTCAAATAACATCAAGTT 909
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 ACTTATGCTGGAACTTCTCAACCAGTTGGATTGGAGCTTCAAATAACATCAAGTT 909

Qy   361 ACTTATGCTGGAACTTCTCAACCAGTTGGATTGGATTGGAGCTTCCAATAAGATCAAGGTT 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 ACTTATGCTGGAACTTCTCAACCAGTTGGATTGGATTGGAGCTTCCAATAAGATCAAGGTT 420

Qy   910 TTGATGGGCCACCAATCGGATTGATATCCTGGATCAAGCCCTCCTTAGACCAGGACGGATA 969
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 TTGATGGGCCACCAATCGGATTGATATCCTGGATCAAGCCCTCCTTAGACCAGGACGGATA 969

Qy   421 TTGATGGCCACCAATCGGATTGATATCCTGGATCAAGACTTCAGACCTGGACGGATT 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 TTGATGGCCACCAATCGGATTGATATCCTGGATCAAGACTTCAGACCTGGACGGATT 480

Qy   970 GACCGGAAAATTGAATTCCAACCCCTAATGAAGAGTCTCGGCTGGATTTTGAAAATC 1029
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 GACCGGAAAATTGAATTCCAACCCCTAATGAAGAGTCTCGGCTGGATTTTGAAAATC 1029

Qy   481 GACAGGAAAATTGAATTCCAAATCCTAATGAAGAGTCGCCGGAGATATCTTGAAAATC 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GACAGGAAAATTGAATTCCAAATCCTAATGAAGAGTCGCCGGAGATATCTTGAAAATC 540

Qy  1030 CATTCTTAGAAGAAGAAGAAATTAATGCGTGGCATTGAATTTGAAGAAGAATTGCCGAGAAGATG 1089
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 CATTCTTAGAAGAAGAAGAAATTAATGCGTGGCATTGAATTTGAAGAAGAATTGCCGAGAAGATG 1089

Qy   541 CATTCAAGAAGAATGAAATTAAATGCGTGGTATTGACTTGAAGAGAGAGATTGCTGAGAAGATG 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 CATTCAAGAAGAATGAAATTAAATGCGTGGTATTGACTTGAAGAGAGAGATTGCTGAGAAGATG 600

Qy  1090 AATGGAGCAATCTGGTGCTGAACTTAAGGCTGTTGCACTGAGAGCTGGAAGTGTTGCTTTG 1149
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 AATGGAGCAATCTGGTGCTGAACTTAAGGCTGTTGCACTGAGAGCTGGAAGTGTTGCTTTG 1149

Qy   601 AATGGAGCATCTGGTGCTGAGCTTAAGGCTGTTGCACTGAGAGCTGGAGATATCTTGAAAATC 660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 AATGGAGCATCTGGTGCTGAGCTTAAGGCTGTTGCACTGAGAGCTGGAGATATCTTGAAAATC 660

Qy  1150 AGGGAGCGCAGGGGTTACACGTGACTC 1174
         ||||||||||||||||||||||||||||
Db  1141 AGGGAGCGCAGGGGTTACACGTGACTC 1174

Qy   661 AGAGAGAGGAGGGGTTACATGTGACAC 685
         ||||||||||||||||||||||||||||
Db   661 AGAGAGAGGAGGGGTTACATGTGACAC 685
```

```
              Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
              Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
              Fujiwake,H., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
              Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
              Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A., and Hayashizaki,Y.
  TITLE       RIKEN integrated sequence analysis (RISA) system-384-format
              sequencing pipeline with 384 multicapillary sequencer
  JOURNAL     Genome Res. 10 (11), 1757-1771 (2000)
  PUBMED      11076861
  MEDLINE     20530913
REFERENCE   4
  AUTHORS     Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
              Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
              Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
              Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
              Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
              Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
              Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
              Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
              Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
              Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
              Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
              Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
              Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
              Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
              Ring,B., Ringwald,M., Rodriguez,B., Sakamoto,N., Sasaki,H.,
              Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
              Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C.A., Wilming,L.,
              Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
              and Hayashizaki,Y.
  TITLE       Functional annotation of a full-length mouse cDNA collection
  JOURNAL     Nature 409 (6821), 685-690 (2001)
  PUBMED      11217851
  MEDLINE     21085660
REFERENCE   5
  AUTHORS     The FANTOM Consortium and the RIKEN Genome Exploration Research
              Group Phase I & II Team.
  TITLE       Analysis of the mouse transcriptome based on functional annotation
              of 60,770 full-length cDNAs
  JOURNAL     Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 1313)
  AUTHORS     Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
              Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
              Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
              Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kawai,M.,
              Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
              Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
              Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
              Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
              Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y.,
              Suzuki,H., Tagami,M., Takahashi,F., Tanaka,T.,
              Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
              Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE       Direct Submission
  JOURNAL     Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
              Physical and Chemical Research (RIKEN), Laboratory for Genome
              Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
              RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
              Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
              URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
              Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
              further details.
              cDNA library was prepared and sequenced in Mouse Genome
              Encyclopedia Project of Genome Exploration Research Group in Riken
              Genomic Sciences Center and Genome Science Laboratory in RIKEN.
              Division of Experimental Animal Research in Riken contributed to
              prepare mouse tissues. First strand cDNA was primed with a primer
              [5'-GAGAGAGAAGCAAGACGTCTTTTTTTTTTTTTTTTTVN 3'], cDNA was
              prepared by using trehalose thermo-activated reverse transcriptase
              and subsequently enriched for full-length by cap-trapper. Second
              strand cDNA was prepared with the primer adapter of sequence [5'
              GAGAGAGAGGATTCTCGGGTNAATTAANTTANTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
```

```
              with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
              Host: SOLR.
FEATURES             Location/Qualifiers
     source          1..1313
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /clone="1313"
                     /db_xref="taxon:10090"
                     /dev_stage="11 days embryo"
                     /tissue_type="whole body"
     misc_feature    39..1259
                     /note="protease (prosome, macropain) 26S subunit, ATPase 5
                     putative"
                     /db_xref="MGI:105047"
                     (MGD|MGI:105047)
     polyA_signal    1294..1299
                     /note=".1299
     polyA_site      1313
                     /note="putative"
                     /note="putative"
     db_xref="MGI:1095485"
     db_xref="FANTOM DB:2700027I02"
     clone="2700027I02"
BASE COUNT      349 a    291 c    389 g    284 t
ORIGIN

Query Match       48.8%;  Score 604;  DB 11;  Length 1313;
Best Local Similarity  71.2%;  Pred. No. 1.5e-149;
Matches  840;  Conservative  0;  Mismatches 335;  Indels  5;  Gaps  3;

Qy    76 AAGCAGGGGCGAGGGCCTCCCGCACTACTATTCTTCAACATTCCACGAGCATCAGTCCTT  135
         |||| |||  |  ||||||| ||| || | |||||||| |||||| ||||||||| ||
Db    81 AAGGCAGGCAGTGGACTCCGTCAATATTATCGTGTCCAATATTATCGTGTCCAGTTGATT  140

Qy   136 CTTCGGCCAAAGACTCATAACCTTCAACCGTCTCCAGGGCTCAGAGAAACGACTTCAATTCT  195
         ||||  |  ||| ||| |  |  |||| | || ||  ||| ||| |||||||||| |
Db   141 GTAAATGATAGGAGCCCAGAATCTCCGTAGACTGCAGGCACAGAGGAATGAGCTGAATGCA  200

Qy   196 AGGGGTGAGGAGATGCTGGGCGAAGAGAATTACAGCTTCTTGCAGGAAACCGCGGTCTGTTGCAG  255
         ||||| || ||  |||  |||| | |||| ||  |||  |||||| || || ||  |
Db   201 AAACTTCGGCCTGTTGCGGGAGGAGCTGCAGCTGGTTGCAGGAACACGGGCTTCCTACGTTGGA  260

Qy   256 GAAGTTGTCAAAGTAATTGGGCCAAGAGACAAAAGTCCTTGTCAAGGTCCACCCCAGAGAGGAAAA  315
         | || ||| | ||| |  ||| |||| || |||||  |||| |||||  ||| ||||
Db   261 GAAGTCGTGAGGGGCCATGGATTAAGAGAAAAGTATTGGTCAAGGTTCCATCCTCTGAGGGCAA  320

Qy   316 TATGTTGTTTGACATTGACAAAAATATTGACATTACAAAAGATTACTCCATCCACTGAGTT  375
         |  | |||||  |||| || ||| | |||  |||||  || || | ||||||  ||| |
Db   321 TTTGTTCGTTGATGTGGATCAGGACAACAATTGATATCAACGATGTGACCCCCAAATTGTCGGGTC  380

Qy   376 GCAACTCCGCAACGACACAGTTAATGTTGTTCTTCA-CTTAGTTCCTGCCAGTAAAAGTTGATCCATT  434
         |  ||  | |||| |  || | | |  | |  ||| || |||| | ||  ||||| |||| |||
Db   381 GCTCTAAGACATGACAAGCTCACACTCTTCACTAAGACTCTTACCTAACACAAGGGTGGGAACCCTTT  440

Qy   435 GGTCAATCTGATGAAAGTTGAGAAAGTTCCGGATTCTTACATATGACATGATTGGTTGGTTT  494
         |||| | ||| |||| | ||| || |||  || | ||||| | |||| ||| |||| |
Db   441 GGTGTCACTAATGATGGTGGAGAAGGTGCCAGAACTCACAACCTACAGGGACGAGTTGCGCGGCCT  500

Qy   495 AGACCCAGCAAATTAAGGAAATTAAAAGGAGGTCATTCAGCACACCAAACACATACCTGAGCT  554
         || |  | ||| ||   | |||  | ||||| | ||| ||| ||| |||  || | |||
Db   501 GGACAACGAGATCAAGGAGAATTAAAGAAGTGATCGGAGCTGGCCCGTGAAGCAGCACCCCGAGCT  560

Qy   555 GTTTGAAAGTCTTGGAATTGCACAACCAAAAGGGGTGTCCTGGCTCCTPATGGGGCCCACTGGTAC  614
         |||| |  |   || |||   | ||| |||| | ||||  ||  | | | |||||| |||| ||
Db   561 CTTTGAAGCACTGGGCATCCCACAGCCCAAAGGGGAGTCCGCTCCTACGGACCCACCAGGCAC  620

Qy   615 AGGTAAAACATTGTTGGCTAGGGCACGTGGCTCATCATACTGACTGTACATTCATCAGGGGT  674
         | || |  | || | |  || ||  | |  | ||| || | |||||  |||| ||||||
Db   621 TGGAGGAACATGTTCGGCCCGGCGGCTCGTGTGGCCAATACTACCTTTATTCATCGTTCGT  680

Qy   675 GTCTGGTTTCTGAGTTAGTTCAGAAAAATTACATTGGAGAAAGTTTCAGAAATGGTCAGGGAAACT  734
         | ||  ||  | |  | |  ||||| |  || | |  ||||| || ||| |||| |||
Db   681 CTCTGGCTCTGGAACTGGTACAGAAAATTCATCACGGGAGGAAGTGGTGTAGGGGAGCT  740
```

```
FEATURES             Location/Qualifiers
     source          1..715
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /cultivar="A17"
                     /db_xref="taxon:3880"
                     /clone="pGESD10M14"
                     /tissue_type="immature seeds"
                     /dev_stage="immature seeds, 11 to 19 days after
                     pollination"
                     /clone_lib="GESD"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Immature seeds, collected from pods ranging in age
                     from 11 to 19 days after pollination, were harvested from
                     greenhouse-grown plants.  Seeds were removed and
                     separated from pod walls and were immediately frozen in
                     liquid nitrogen.  Seeds throughout the age range were
                     pooled for mRNA extraction.  cDNA was prepared from polyA+
                     enriched RNA.  The cDNA was directionally ligated into
                     the UniZap XR vector from Stratagene and packaged using
                     Gigapack III Gold packaging extracts.  Plasmids containing
                     cDNA inserts were excised from the recombinant lambda-Zap
                     phage using Ex-assist helper phage and propogated in
                     XLOLR cells."
BASE COUNT      213 a    117 c    188 g    197 t
ORIGIN
```

```
Query Match             48.1%;  Score 602.8;  DB 12;  Length 715;
Best Local Similarity   92.3%;  Pred. No. 2.4e-149;
Matches  657;  Conservative  0;  Mismatches 52;  Indels  3;  Gaps  2;
```

```
Qy   478 GACATGATTGGTGGTTTAGACCAGCAAATTAAAGAAATAAAAGAAGGTCA--TTGAGCTAC 535
         ||||||||||||| ||||||| ||||||| ||||||| |||||||||||||   ||||||| ||
Db     5 GGCATGATTGGTGTTTAGACAACAAATCAAAGAAGNAATAAAAGAGGTCNTTTTGAGCTTC 64

Qy   536 CAATCAAACATCCTGAGCTGTTTGAAGTCTTGGAATTGCACAACCAAAAGGGTGTCCTGC 595
         |||| |||||||||||| |||| ||||| |||| ||| ||||||||||||||| || |||
Db    65 CAATTAAACATCCTGAGTTGTTCGAAAGTCTTCGAAGTTCTCAACCAAAAGGGTCTTTGC 124

Qy   596 TCTATGGGCGCCACCTGGTACAGGTAAAACATTGTTGGCTAGGGGCAGTGGCCTCATCATACTG 655
         | ||||||| |||||||| |||| |||||||||||||||||| |||||| |||| ||| |||||
Db   125 TTTATGGGCCCACCTGGTCACAGGAAAAACATTGTTGGCTAGGGCCAGTGAGCACATCACTACTG 184

Qy   656 ACTGTACATTCACACGGTGTCTCGGTTCTGAGTTAGTTCAGAAATACATTGGAGAAGGTT 715
         | |||||||||| |||||||| ||||||| |||| |||||||||||||| | ||||||||
Db   185 ATTGTACATTCATCAGGGTGTCTGGTTCTGGGTTGTTCAGAAATACATCCGGAGAAGGTT 244

Qy   716 CTAGAATAGGTCAGGGAACTTTTTTGTTATGGGCAACAGATGCTCCATCAATTATATTCA 775
         ||||||||| | ||||||||||| ||||||||||| ||| ||| |||||||||| | |||
Db   245 CTAGAATAGTTAGGGAACTTTTTGTTATGGGCAGGAGACATGCCCCATCAATTATCTTCA 304

Qy   776 TGGATGAAAATTGACAGTATTGGATCTCGCTCGGATGGAATCTCGGAACGTGGCAACGCGTGATA 835
         ||||| |||||| |||||| |||| ||||||||| |||||||||||||| | ||||| | |||||
Db   305 TGGACGAAAATGACAGTATTGGANCTCGCTCGTATGGAATCTCGGAAGTCGGCAATGGTCATA 364

Qy   836 GTGAGGTACAGCGTACTATGCTGGAACTTCTCAACCAGTGGATTGAATTGAAGCTTCAA 895
         |||| || |||||||||||||||||||||||||||||||| ||     |||||||||||
Db   365 GTGAAGTGCAGCGTACTATGCTGGAACTTCTCAACCAGTGGATCAAGCCCTCCTTA 424

Qy   896 ATAAGATCAAGGTTTTGATGGTCCACCAATCGGATTGATATCCTGGATCAAGCGCCGCGAG 955
         ||||||||||||||||||||| |||||||||||||||||||||||||||||||| | ||
Db   425 ATAAGATCAAGGTTTTGATGGCCACCAATCGGATTGATATCCTGGATCAAGCGACTCTGA 484

Qy   956 GACCAGGACGGATAGACCGGAAAAATTGAATTCCAACCCCTAATGAAGAAGAGATCTCGGCTGG 1015
         |||||||| ||||||| |||||||||||||||||||| |||||||||||||||| | ||||| ||
Db   485 GACCTGGACGGATGACAGGAAAAATTGAATTTCCAAATCCTAATGAAGAAGAGTCTCGGCCTGG 544

Qy  1016 ATATTTTGAAAAATCCATTCTAGAGAAATGAAATTTAATACCGTGGCATTGATTTGAAGAGAGA 1075
         |||||||||||||| ||| || || ||||||| |||||||| || ||| |||||| |||||||
Db   545 ATATCTTGAAAAATCCATTCCAGAGAAATCCATTAATGCGGTGGTATTGAAGCTGTTGCACTGAGA 604

Qy  1076 TTGCCGAGGAAGACGAAATGGAGCAGCATCCTGTCGGTGAAGCTGTTTGCAGTGAGGCTG 1135
         ||||| |||||||||||||| ||||||||||| || |||||||||||||| |||||||||
Db   605 TTGCTGAGAGATGAATGGGACATCCTGGTGCTGAGCTTAAGGCTGTGTGCACTGAAGTG 664

Qy  1136 GAATGTTTTGGCTTTGA-GAAGAGAGGGTACACAGAGGGGATTTTGA 1187
         |||||| ||||||||| | ||||||||||||| |||||||||||||
Db   665 GAATGTTGGCCTTTGA-GAAGAGAGGGTACAATGGACAACAGAGGATTTTGA 715
```

```
Qy   735 TTTGTTAGGGCAGGGAACATGCTCCATCAATTACTTCATGGATGAAATTGACAGTAT 794
         ||||| || | ||| |||||||||||||| |||| |||||| |||||||||||| |||
Db   741 GTTTGTCATGGCTCCCGAGACATGCTCCATCCATCATCTTCATCGACGGATTGACTCTAT 800

Qy   795 TGGATCTGCTCGGATGGAATCTGAAGTGGCAACGGGTGATAGTAGTAGGTACACCGTACTAT 854
         ||| ||| ||| ||| || |     ||| ||||| ||||| || || ||| ||| | ||||
Db   801 TGGGCTCCTCACGGGCTGGAGGGGCCTCTCGG---AGGCGACAGTGAGGGTAGCAGCGCACGAT 857

Qy   855 GCTCGGAACTTCTCAACCAGTTGCATTGGATTGGAAGCTTCAAATAAGATAAGGTTTTGAT 914
         |||||||||| ||||| |||| | |||| ||| ||||| |||||| |||| |||||| ||
Db   858 GCTCGGAACTGCTCAATCAGCTGGATGGCTTGAGGCCACCAAGAAATATCAAGGTTATCAT 917

Qy   915 GGCCACCAATCGGATTGATATCCTTAGACCAGGACGGATAGACCG 974
         ||||||||| |||||||||||||| |||| |||| |||||| ||
Db   918 GGCCTACTAATAGGATTGATATCCTGCGACTCTGCCCTGTTCCTTCCTCGGGACGGATTGACAG 977

Qy   975 GAAAATTGAATTCCAACCCCTAATGAAGAAGAGATTTTTGAAAAATCCATTC 1034
         |||||||||||||| ||||| || ||||| | | |||||||||||||||| ||
Db   978 AAAAATTGAATTCCACCCCCAACGAAGCCCCCGGCCTGGGACATTTTGAAAAATCCACTC 1037

Qy  1035 TAGAGAAATGAAATTTAATACCGTGGCATTGATTTGAAGANAGATTGCCCGAGAAGGAATGG 1094
         | |||||||||| || || |||||| || | ||||||| ||| ||||| |||| |||||||
Db  1038 TCGGAAAAATGAAACTTGACCAGGGGATCAAACCTGAGGAAAATTGCTCAGTGTGATNGCCGGG 1097

Qy  1095 AGCATCTGGTGCTGAACTTAAGCCGTGTTTTGCACTGAACCTTGAAATGTTTGCTTTGAGGGA 1154
         ||| |||| | ||||| | |||| || | | | ||| ||||| | ||||| ||| | ||||||
Db  1098 AGCTTCAGGGCTGAGTGGAGGGGTGTGTGCCACAGAACGTGGCATG-TACCCCTGCGGGA 1156

Qy  1155 GCGGAGGGTACACCTGACTCAGGAGGATTTTGAGATGGTCCGTGGGCGAAGGGTGATGGAAAAA 1214
         || | || || || ||||||||||||| ||||||||| | || |||||| ||||| ||| ||||
Db  1157 ACGGGCGTGTCCACGTCACTCAGGAGGACTTTGAGATGGGCAGTAGCTAAGGTCATGCAAAA 1216

Qy  1215 GGAGACTGAAAAAAACATGTCATTGCGGAAGGTTGTGGAG 1254
         |||| | | |||||||||||| ||| ||||| ||| || ||
Db  1217 GGATAGTAGACGAAAAACATGTCCATCAAGAAGCTATGGAAG 1256
```

```
RESULT 11
BI311279
```

```
LOCUS       BI311279                715 bp    mRNA    linear   EST 20-JUL-2001
DEFINITION  EST531313029 GESD Medicago truncatula cDNA clone pGESD10M14 5' end,
            mRNA sequence.
ACCESSION   BI311279
VERSION     BI311279.1  GI:14985606
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 715)
  AUTHORS   Grusak,M.A., Samac,D.A., Town,C.D., Van Aken,S., Utterback,T.,. Cho
            ,J. and Fraser,C.M.
  TITLE     ESTs from developing reproductive tissues of Medicago truncatula
  JOURNAL   Unpublished
COMMENT     Contact: Michael A. Grusak
            USDA/ARS Children's Nutrition Research Center
            Baylor College of Medicine
            1100 Bates Street, Houston, TX 77030-2600, USA
            Tel: 713-798-7044
            Fax: 713-798-7078
            Email: mgrusak@bcm.tmc.edu
            B398134e
            TIGR sequence name: MTPAR79TK
            More information is available at: www.medicago.org
            Seq primer: Skmod (CTA gAA CTA gtg gAT CC),
```

RESULT 12

```
LOCUS       BM308101     591 bp    mRNA    linear    EST 02-JAN-2002
DEFINITION  sak41g08.y1 Gm-c1036 Glycine max cDNA clone SOYBEAN CLONE ID:
            Gm-c1036-5703 5' similar to TR:Q9SCD6 Q9SCD6 26S PROTEASOME SUBUNIT
            8. ; mRNA sequence.
ACCESSION   BM308101
VERSION     BM308101.1  GI:18039807
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1 (bases 1 to 591)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact: ccu@resgen.com web site:
            www.resgen.com
            Seq primer: -40RP from Gibco
            High quality sequence stop: 430.
FEATURES             Location/Qualifiers
     source          1..591
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /db_xref="taxon:3847"
                     /clone="SOYBEAN CLONE ID: Gm-c1036-5703"
                     /tissue_type="somatic embryos cultured on MSD 20"
                     /clone_lib="Gm-c1036"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This
                     cDNA library was constructed from mRNA isolated from
                     somatic embryos (age ranging from 2 months to 9 months)
                     cultured on MSD 20. The library was prepared using the
                     Life Technologies pSuperScript cDNA library construction
                     kit. Complementary DNA was synthesized from mRNA using a
                     poly (dT) sequence with a NotI restriction site. SalI
                     linkers adapters were ligated to the blunt-ended cDNA
                     fragments followed by NotI digestion. The cDNA fragments
                     were directionally cloned into the NotI-SalI restriction
                     site of the pSPORT1 vector. The ligated cDNA fragments
                     were transformed into E.coli ElectroMax DH10B host cells.
                     This library was constructed in the laboratory of Dr. Lila
                     Vodkin by Anu Khanna at the University of Illinois at
                     Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"
BASE COUNT      177 a    113 c    136 g    165 t
ORIGIN

  Query Match             47.1%;  Score 591;  DB 12;  Length 591;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-146;
  Matches   591;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   187 CTCAATTCTAGGGTGAGGATCCTCGCGGAAGAATTACAGCTTCTGCAGGAACCCGGCTCT 246
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CTCAATTCTAGGGTGAGGATGCCTCGCGGAAGAATTACAGCTTCTGCAGGAACCCGGCTCT 60
```

RESULT 13

```
LOCUS       BI945194     699 bp    mRNA    linear    EST 30-NOV-2001
DEFINITION  sb32ell.y1 Gm-c1012 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1012-165 5' similar to TR:Q9SCD6 Q9SCD6 26S PROTEASOME SUBUNIT
            8. ; mRNA sequence.
ACCESSION   BI945194
VERSION     BI945194.1  GI:16282766
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1 (bases 1 to 699)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            Seq primer: -40RP from Gibco

Qy   247 TATGTCCGGTGAAGTTGTCAAAGTAATGGGCAAGAACAAAAGTCCTTGTCAAGGTCCACCA 306
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TATGTCCGGTGAAGTTGTCAAAGTAATGGGCAAGAACAAAAGTCCTTGTCAAGGTCCACCA 120

Qy   307 GAAGGAAAATATGTTGTTGACATTGACAAAAATATTGACATTACAAAGAATTACTCCATCC 366
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GAAGGAAAATATGTTGTTGACATTGACAAAAATATTGACATTACAAAGAATTACTCCATCC 180

Qy   367 ACTAGAGTTGCACTCCGCAACGACACAGTTATGTTCTTCACTTAGTTCTCGCCAAGTAAGT 426
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 ACTAGAGTTGCACTCCGCAACGACACAGTTATGTTCTTCACTTAGTTCTCGCCAAGTAAGT 240

Qy   427 GATCCATTGGTCAATCTGATGAAAGTTGAGAAAGTTGAGAAAGTTCCGGATTCTACATATGACATT 486
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GATCCATTGGTCAATCTGATGAAAGTTGAGAAAGTTCCGGATTCTACATATGACATGATT 300

Qy   487 GGTGGTTTAGACCCAGCAAATTAAAGAAAATAAAAGAGGGTCATTGAGCTACCAATCAAACAT 546
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GGTGGTTTAGACCAGCAAATTAAAGAAAATAAAAGAGGGTCATTGAGCTACCAATCAAACAT 360

Qy   547 CCTGAGCTGTTTGAAAGTCTTGGAATTGCACACACCAAAGGGGTGTCCTGGCTTCATGGCCA 606
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CCTGAGCTGTTTGAAAGTCTTGGAATTGCACACACCAAAGGGGTGTCCTGGCTTCATGGGCCA 420

Qy   607 CCTGGTACAGGTAAAACAATTGTTCGGCCAGGGACACATGCCTCCATCAATTATCTTCATG 666
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CCTGGTACAGGTAAAACATTGTTGGCTAGGGACAGTGGGCTCATCATACGACTGTACATTC 480

Qy   667 ATCAGGGGTGTCTGGGTTCTGAGTTAGTTCAGAAAATACATTGGAGAAGGTTCTAGAATGGTC 726
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ATCAGGGGTGTCTGGTTCTGAGTTAGTTCAGAAAATACATTGGAGAAGGTTCTAGAAATGGTC 540

Qy   727 AGGGAACTTTTTGTTTTGTTTATTGGCCAGGGGAACATGCCTCCATCAATTATCTTCATG 777
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 AGGGAACTTTTTGTTATGGCCAGGGGAACATGCCTCCATCAATTATCTTCATG 591
```

High quality sequence stop: 425.

```
FEATURES             Location/Qualifiers
     source          1..699
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1012-165"
                     /db_xref="taxon:3847"
                     /tissue_type="Apical shoot tips, 9-10 day old etiolated
                     seedlings"
                     /clone_lib="Gm-c1012"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II XR; Site 1: EcoRI; Site 2:
                     XhoI. This cDNA library was constructed from mRNA isolated
                     from the apical shoots of 9 to 10 day old etiolated
                     seedlings. The shoot tips including any emerged leaves
                     were harvested for mRNA isolation. The cDNA library was
                     prepared using the Stratagene pBluescript II XR cDNA
                     library construction kit. Complementary DNA was
                     synthesized from mRNA using a primer consisting of a poly
                     (dT) sequence with a XhoI restriction site. EcoRI adapters
                     were ligated to the blunt-ended cDNA fragments followed by
                     XhoI digestion. The cDNA fragments were directionally
                     cloned into the EcoRI-XhoI restriction site of the
                     pBluescript vector. The ligated cDNA fragments were
                     transformed into XL10-Gold host cells. This library was
                     constructed by Dr. Randy Shoemaker and Dr. John
                     Expelding."
BASE COUNT      208 a    148 c    151 g    183 t      9 others
ORIGIN
```

Query Match 46.9%; Score 587.8; DB 12; Length 699;
Best Local Similarity 92.6%;
Matches 646; Conservative 0; Mismatches 46; Indels 6; Gaps 3;

```
Qy    95 GCCACTACTATTCTTCTCAACATCACAGAGCATCAGCTCCTTCTTCCGCCAAAAGACTCATA 154
         |||||||||||||||||||||||| ||||||| ||||||||||||| |||||||||||||||
Db     8 GCCACTACTATTCTTCTCAACATCCACGAGCATCCAGCTCCTTCTTCGGCCAAAAGACTCATA 67

Qy   155 ACCTCAACCGTCCGAGGCTCAGAGAAAACGACCTCAATTCTAGGTGGGTGGAGGAATGCTGCGCG 214
         ||||||||||| ||||||||||||||||||||||||||||||| ||||||||||| ||||||||
Db    68 ACCTCAACCGTCTCGAGGCTCAGAGAAAACGACCTCAATTCTAGGGTGGGTGGAGGATGCTGCGCG 127

Qy   215 AAGAATTACAGCTTCTGCAGGACACCCGGCGCTCTTATGTCGGTGAAGTTGTCAAAGTAATGG 274
         |||||||||||||||||||||| ||||||||||||| ||||||||||||||||||||||||||
Db   128 AAGAATTACAGCTTCTGCGAGGAACCCGGCGCTCTTCTTATGTCGGTGAAGTTGTCAAAGTAATGG 187

Qy   275 GCAAGAACAAAAGTCCTTGTTCAAGGTCACCCAGAAGAGGAAAATATGTTGTTGACATTGACA 334
         |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   188 GCAAGAACAAAAGTCCTTGTTCAAGGTCACCCAGAAGAGGAAAATAGTGTTGTTGACATTGACA 247

Qy   335 AAAATATTGACATTACAAAAGATTACTCCATCCACTAGAGTTGCCACTCCGCACGACAGTT 394
         ||||||||||||| |||||||||||||||||||||||||||||||||||| |||||||||||
Db   248 AAAATATTGACATTACAAAAGATTACTCCATCCACTAGAGTTGCCACTCCGCACGACAGTT 307

Qy   395 ATGTTCTTTCACTTAGTTCTGCCAGTAAAGTTGGTCAATCTGATGAAAGTTG 454
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   308 ATGTTCTTTCACTTAGTTCTGCCAGTAAAGTTGGTCAATCTGATGAAAGTTG 367

Qy   455 AGAAAGTTCCCGATTCTACATATGACATGATTGGTGGTGTTTAGACCAGCAAATTAAGAGA 514
         |||||||| ||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   368 AGAAAGTTCCCGATTCTACATATGACATGATTGGTGGTGTTTAGACCAGCAAATTAAGAGA 427

Qy   515 TAAAAGAGGTCATTGAGCTACCAAATCAAACATCCTGAGCTGTTTGAAAGTCTTGGAATTG 574
         ||||||||||||||| || ||||| || ||||||||||||||||||||||||||||||||
Db   428 TAAAAGAGGTCATTGAGCTACCAATCAAACATCCTGAGCTGTTTGAAAGTCTTGGAATTG 487

Qy   575 CACAACCAAAGGGTGTCTGCTTGTACTGCCTCGCCCTGGTACAGGTAAAACAATTGTTGGCTA 634
         |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   488 CACAACCAAAGGGTCATCANTACTGCTGTACATTCATCAGGGCCACCTGGTTACAGGTAAAACAATTGTTGGGCT 547

Qy   635 GGGCAGTGGCTCATCATACTGACCTGTACATTCATTCAGGGTGTCTGGTTCTGAGTTAGTTC 694
         ||||||||||||||||| |||||||||||| |||||||||||||||||||||||||| |||
Db   548 AGGCAGTGGCTCATCANTACTGCTGTGACATTTCATCANGGGTGTCTGGTTGTTGAGTTAGTAGTTAG 607
```

```
Qy   695 AGAAATACATTGGAGAAGGTTCTTAGAAATGGTCAGGGGAACTTTTTGTTATGGCCAGGGAAC 754
         |||||||  ||| ||| ||||| |||||| ||||||| ||||||||  ||||||||||  |||
Db   608 ANATACA---TTGNAAANGGTCTAGAAATGTCAGGGAACCTTTT--GTATGGCCAGNNGAC 662

Qy   755 ATGCTCCATCAATTATCTTCATGGATGRAAATTGACAGT 792
         |||||| ||| |||||||| |||||||| ||||||||||
Db   663 ATGCTGCATC-ATTATCTNCATGGATGAAAATTGACAGT 699
```

RESULT 14
```
LOCUS       BI944357
BI944357                 627 bp    mRNA    linear   EST 28-NOV-2001
DEFINITION  saa61b06.y1 Gm-c1060 Glycine max cDNA clone GENOME SYSTEMS CLONE
            ID: Gm-c1060-1211 5' similar to TR:Q9SCD6 Q9SCD6 26S PROTEASOME
            SUBUNIT 8. ;, mRNA sequence.
ACCESSION   BI944357
VERSION     BI944357.1  GI:16281184
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
            eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 627)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Brpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            Seq primer: -40RP from Gibco
            High quality sequence stop: 409.
FEATURES             Location/Qualifiers
     source          1..627
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1060-1211"
                     /tissue_type="Root, 2 week seedlings"
                     /clone_lib="Gm-c1060"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
                     XhoI. The cDNA library was constructed from mRNA isolated
                     from root tissue of 2 week old seedlings for PI468916.
                     Complementary DNA was synthesized from mRNA using a primer
                     consisting of a poly(dT) sequence with a XhoI restriction
                     site. EcoRI adapters were ligated by XhoI digestion. The cDNA
                     fragments were directionally cloned into the EcoRI-XhoI restriction
                     site of the pBluescript vector. The ligated cDNA fragments
                     were transformed into DH10B host cells (Gibco BRL). This
                     library was constructed in the laboratory of Dr. Randy
                     Shoemaker at Iowa state university."
BASE COUNT      187 a    116 c    146 g    177 t      1 others
ORIGIN
```

Query Match 46.9%; Score 587.6; DB 12; Length 627;
Best Local Similarity 96.0%;
Matches 602; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

RESULT 15

```
LOCUS       AW458284                632 bp    mRNA    linear    EST 03-DEC-2001
DEFINITION  sh80g08.y1 Gm-c1016 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1016-6519 5' similar to SW:PRS8_MOUSE P52915 26S PROTEASE
            REGULATORY SUBUNIT 8 ;, mRNA sequence.
ACCESSION   AW458284
VERSION     AW458284.1  GI:7028501
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 632)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk
            ,R., Ritter,B., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished
  COMMENT   Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 for further information
            Call: (800)-533-4363 or contact via email: ccueresgen.com
            Insert Length: 1727   Std Error: 0.00
            Seq primer: -40RP from Gibco
            High Quality Sequence stop: 422.
FEATURES             Location/Qualifiers
     source          1..632
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1016-6519"
                     /tissue_type="immature flowers of field grown plants"
                     /lab_host="XL10-Gold"
                     /clone_lib="Gm-c1016"
                     /note="Vector: pBluescript II XR; Site 1: EcoRI; Site 2:
                     XhoI; This cDNA library was constructed from mRNA isolated
                     from immature flowers of field grown plants. The cDNA
                     library was prepared using the Stratagene pBluescript II
                     XR library construction kit. Complementary DNA was
                     synthesized from mRNA using a primer consisting of a poly
                     (dT) sequence with a XhoI restriction site. EcoRI adapters
                     were ligated to the blunt-ended cDNA fragments followed by
                     XhoI digestion. The cDNA fragments were directionaly
                     cloned into the EcoRI-XhoI restriction site of the
                     pBluescript vector. The ligated cDNA fragments were
                     transformed into XL10-Gold host cells. This library was
                     constructed by Dr. Randy Shoemaker and Dr. John
                     Erpelding."
BASE COUNT     183 a    146 c    148 g    154 t    1 others
ORIGIN
```

Query Match   46.3%;   Score 580.8;   DB 9;   Length 632;
Best Local Similarity   95.5%;   Pred. No. 1.6e-143;
Matches 597;   Conservative   0;   Mismatches 28;   Indels   0;   Gaps   0;

```
Qy          24 GAAGCAATGCGCGGGAGGCGTTACCCGGAGGGCGGAATTGCTCCGGCCAAGCCCACCAAGCCAGGG      83
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           8 GAAGCAATGCGCGGGAGGCGTTACCCGGAGGGCGGAATTGCTCCGGCCAAGCCCACCAAGCCAGGG      67

Qy          84 CGAGGGCGCCTCCGCCACTACTATTTCTCTCAACATCCACGACCATCAGCTCTCTTCTTCGGCCA     143
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db          68 CGAGGGCGCCTCCGCCACTACTATTTCTCTCAACATCCACGACCATCAGCTCTCTTCTTCGGCTCA     127

Qy         144 AAAGACTCATAACCTCCAACCGTCTCGGAGGCTCAGAGAACGACCTCAATTCTAGGGTGAG      203
               |||  |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db         128 AAAGACGCATAACCTCAACCGTCTCGGAGGCTCAGAGAAACGACGTCAATTCCAGGGTGAG      187

Qy         204 GATGCTGCGCGGAAGAATTACAGCTTCTGCAGGAACCCGGCCTCTTATGTCGGGTGAAGTTGCT      263
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         188 GATGCTGCGCGGAAGAATTACAGCTTTTGCAGGGAACCAGGCTCTTATGTGGGTGAAGTTGCT      247

Qy         264 CAAAGTAAATGGGCAAGACAAAAGTCCTTGTCAAGGTCCACCCAGAAGGAAAAATATGTTGT      323
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         248 CAAAGTCATGGGCAAGACAAAAGTCCTTGTCAAGGTCCACCCAGAAGGAAAAATATGTTGCT      307

Qy         324 TGACATTGACAAAAATTGACATTACAAAGATTACTCCATCCACTAGAGTTGACTCCG      383
               ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         308 TGATATTGACAAAAGTATTGACATTACAAAGATTACTCCATCCACTAGAGTTGCACTTCG      367

Qy         384 CAACGACAGTTATGTTTCTTCACTTAGTTCTGCCAAGTAAAGTTGATCCATTGGTCAATCT      443
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         368 CAACGACAGTTATGTTTCTTCACTTAGTTCTGCCAAGTAAAGTTGATCCATTGGTCAATCT      427

Qy         444 GATGAAAGTTGAGAAAGTTCCCGGAATTCTACATATTGGTGGGTTAGACCAGCA      503
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         428 GATGAAAGTTGAGAAAGGTCCCGGATTCTACTTAGGACATGATTGGTGTGGTTTAGACCAGCA      487

Qy         504 AATTAAAGAAATAAAAAGAGGTCATTGAGCTACCAATCAAACATCCTGAGCTGTTTGAAAG      563
               ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         504 ...
```

(alignment continues)

```
Qy         607 CCTGGTACAGGTAAAACAATTGTTGTTGGGCTAGGGCGTCATCATACTACTGACTGTCATTC      666
Db         666

Qy         547 CCTGAGCTGTTTGAAAGTTCTTGAATTCGACACAGCCAAAGGGGTGTCTTCTGCCTCTATGGGCA      606
Db         606

Qy         487 GGTGGTTTAGACCAGCAGAAAATTAAAGAAATAAAAAGAGGTCATTGAGCTACCAATCAAACAT      546
Db         546

Qy         427 GATCCATTGGTCAATCTGATGAAAGTTGAGAAAGTTCCCGGAATTCTACATATGACATGATT      486
Db         486

Qy         367 ACTAGAGTTGCACTCCGCAACGACAGTTATGTTTCTTCACTTAGTTCTGCCAAGTAAAGTT      426
Db         426

Qy         307 GAAGGAAAAATATGTTGCTTGATATTGACATTACAAAGATTACAAAGATTACTCCATCC      366
Db         366

Qy         247 TATGTCGGGTGAAGTTGCTCAAAGTAATGGGCAAGACAAAAGTCCTTGTCAAGGTCCACCCA      306
Db         306

Qy         187 CTCAATTCTAGGGTGAGGATGCTCGCGCGGAAGAAGTTACAGCTTCTGCAGGAACCCGGCTCT      246
Db           1 CTCAATTCCAGGGTGAGGATGCTGCGGCGGAAGAATTACAGCTTTTGCAGGAACCAGGGCTCT      60

Qy         727 AGGGAACTTTTGTTATGGCCAGGAACATGCTCCATCCATCCAATTTATCTTACTTCATTGGTGATGAAATT      786
Db         786

Qy         481 ATCAGGGTGCTGCGTTCTCGAATTAGTTCATAAATACAATTGGAGAAGGTTCTAGAATGGTC      540
Db         540

Qy         667 ATCAGGGTGCTGCGTTCTCGGTTAGTTCAGAAATACAATCATTGGAGAGAGGTTCTAGAATGGTC      726
Db         726

Qy         787 GACAGTATTGGATCTGCTCCGATGGAA      813
Db         601 GACAGGATTGGATCTGCTCGGATTGAA      627
```

Db   488 NAITAAAGAAAATAAAAGAATGTCAITGAGTTACCAATCAAAACATCCTGAGCTGTTTGACAG 547
         |||||||||||||||| ||||||||| |||||||||||||||||||||||||||||||||||| ||
Qy   564 TCTTGGAATTGCACAACCAAAGGGTGTCCTGCTTCTATGGGCCACCTGGTACAGGTAAAAC 623
         ||||||||| ||||| | ||||||||||| ||||||||||||||||||||||| |||||||
Db   548 TCTTGGAATTGCACTACCAAATGGTGTGTCCTGCTCTCTATGGGCCACCTGGTACAGGTAAAAC 607

Qy   624 ATTGTTTGGCTAGGGCCAGTGGCTCAT 648
         |||||||||||| ||||||||||||||
Db   608 ATTGTTTGGCTATGGCAGTGGCTCAT 632

Search completed: November 9, 2003, 11:15:00
Job time : 3031 secs

QM protein - protein search, using sw model

Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.6

Run on: November 9, 2003, 11:15:06 ; Search time 81 Seconds
                (without alignments)
                819.108 Million cell updates/sec

Title: US-09-462-972-2
Perfect score: 2116
Sequence: 1 MALVGVELKHAAEGVPEANC..........VAKVWKKETEKNMSLRKLMK 418

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   A_Geneseq 19Jun03:*
     1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
     2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
     3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
     4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
     5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
     6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
     7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
     8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
     9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
    10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
    11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
    12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
    13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
    14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
    15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
    16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
    17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
    18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
    19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
    20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
    21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
    22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
    23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
    24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2116	100.0	418	20	AAW97650	Soybean SUG1 polyp
2	1905.5	90.1	419	21	AAG24052	Arabidopsis thalia
3	1904	90.0	422	20	AAW97652	Corn SUG1 polypept
4	1891	89.4	405	21	AAG24053	Arabidopsis thalia
5	1860	87.9	382	20	AAW97651	Corn SUG1 polypept
6	1740	82.2	350	21	AAG24054	Arabidopsis thalia
7	1670.5	78.9	402	21	ABP71488	Human cancer assoc
8	1670.5	78.9	402	23	ABP41940	Human ovarian anti
9	1665	78.7	406	20	AAY27553	TR-interacting JL1
10	1665	78.7	406	20	AAW92374	Human JL1 protein.
11	1665	78.7	406	20	AAW85089	Thyroid hormone re
12	1660.5	78.5	406	22	ABB58177	Drosophila melanog
13	1655	78.2	406	17	AAW06100	Human 26S proteaso
14	1654	78.2	406	15	AARS3528	Thyroid hormone re
15	1622.5	76.7	399	22	ABB58875	Drosophila melano
16	1540.5	72.8	401	23	ABP71488	Candida albicans e
17	1508.5	71.3	389	24	ABJ25711	Aspergillus fumiga
18	1507.5	71.2	389	24	ABJ26311	Aspergillus fumiga
19	1501	70.9	321	22	AAG28325	Arabidopsis thalia
20	1497.5	70.8	405	20	AAR53529	SUG1 transcription
21	1497.5	70.8	405	20	AAY27554	SUG1 transcription
22	1497.5	70.8	405	15	AAW92375	S. cerevisiae SUG1
23	1497.5	70.8	299	21	AAG28326	Arabidopsis thalia
24	1394	65.9	299	21	AAG28327	Arabidopsis thalia
25	1105	52.2	241	20	AAR94600	Putative P. abyssi
26	937.5	44.3	399	17	AAR94600	S4 protein. Homo
27	901	42.6	440	23	ABB57208	Mouse ischaemic co
28	901	42.6	439	22	ABB61378	Drosophila melanog
29	895	42.3	439	22	AAG52876	Arabidopsis thalia
30	894.5	42.3	426	21	AAG52876	Arabidopsis thalia
31	886.5	41.9	348	21	AAG06967	Arabidopsis thalia
32	886.5	41.9	348	21	AAG52877	Arabidopsis thalia
33	884	41.8	433	22	ABB58036	Drosophila melanog
34	883	41.7	347	21	AAG06968	Arabidopsis thalia
35	883	41.7	347	21	AAG52878	Arabidopsis thalia
36	877	41.4	389	18	AAW37505	Human 26S proteaso
37	874	41.3	389	21	AAW37505	Human 26S proteaso
38	869.5	41.1	451	24	ABJ26363	Aspergillus fumiga
39	869	41.1	397	22	ABJ26198	Aspergillus fumiga
40	868.5	41.0	463	24	ABJ25763	Drosophila melano
41	868.5	41.0	393	24	AAG46306	Aspergillus fumiga
42	862	40.7	399	21	AAG52881	Arabidopsis thalia
43	862	40.7	501	21	AAG52881	Arabidopsis thalia
44	860	40.6	519	21	AAG52880	Arabidopsis thalia
45	860	40.6	598	21	AAG52879	Arabidopsis thalia

RESULT 1
AAW97650

ID   AAW97650 standard; Protein; 418 AA.
XX
AC   AAW97650;
XX
DT   10-MAY-1999 (first entry)
XX
DE   Soybean SUG1 polypeptide.
XX
KW   SUG1; transcription factor; coactivator; soybean; 26S proteasome.
XX
OS   Glycine max.
XX
PN   WO9902689-A1.
XX
PD   21-JAN-1999.
XX
PF   07-JUL-1998; 98NO-US13992.
XX
PR   11-JUL-1997; 97US-0893401.
XX
PA   (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI   Allen SM, Odell JT;
XX
DR   WPI; 1999-120890/10.
DR   N-PSDB; AAX07039.
XX
PT   Newly isolated nucleic acid fragment encoding a plant SUG1 protein
PT   homologue - useful for regulating expression of specific genes
PT   normally controlled by SUG1, and for screening for plant SUG1

ALIGNMENTS

```
PT   protein inhibitors
XX
PS   Claim 1; Page 25-26; 52pp; English.
XX
CC   This polypeptide comprises the soybean homologue of mouse SUG1
CC   transcriptional coactivator. Its amino acid sequence was deduced
CC   from cDNA clone sel.pk0023.b5 (see AAX07039), which was isolated
CC   on the basis of homology to mouse, yeast and Xenopus SUG1 ESTs.
CC   SUG1 is involved in regulation of gene expression and also functions
CC   as a regulatory component of the 26S proteasome, and hence is
CC   involved in the regulation of protein turnover. Soybean, corn,
CC   wheat and rice SUG1 polynucleotides (see AAX07039-43) and polypeptides
CC   (see AAW97650-54) are provided, as well as chimeric genes in which the
CC   SUG1 polynucleotide is in sense or antisense orientation, and
CC   expression results in production of altered levels of SUG1 protein
CC   in transformed host cells. SUG1 proteins can be used to modulate
CC   expression of specific genes whose promoters are normally regulated
CC   by SUG1 or targeted by transcription factors that normally interact
CC   with SUG1. Host cells can be used to screen for compounds that
CC   modulate SUG1 activity.
XX
SQ   Sequence   418 AA;

Query Match             100.0%;  Score 2116;  DB 20;  Length 418;
Best Local Similarity   100.0%;  Pred. No. 2.5e-175;
Matches  418;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1  MALVGVELKHAAEGVPEANCSAKPTKQGEGLRHYSLNIHEHQLLLRQKTHNLNRLEAQR   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  MALVGVELKHAAEGVPEANCSAKPTKQGEGLRHYSLNIHEHQLLLRQKTHNLNRLEAQR   60

Qy    61  NDLNSRVRMLREELQLLQEPGSYVGVEVKVKMGKNKVLVKVHPEGKYVVVDIDKNIDITKIT  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  NDLNSRVRMLREELQLLQEPGSYVGVEVKVKMGKNKVLVKVHPEGKYVVVDIDKNIDITKIT  120

Qy   121  PSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQIKEIKEVIELPI  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  PSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQIKEIKEVIELPI  180

Qy   181  KHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKYIGEGSR  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  KHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKYIGEGSR  240

Qy   241  MVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNK  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  MVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNK  300

Qy   301  IKVLMATNRIDILDQALLRPGRIDRKIEFPTPNEESRLDILKIHSRRMNLMRGIDLKKIA  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  IKVLMATNRIDILDQALLRPGRIDRKIEFPTPNEESRLDILKIHSRRMNLMRGIDLKKIA  360

Qy   361  EKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK  418
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  EKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK  418


RESULT 2
AAG24052
ID   AAG24052 standard; Protein; 419 AA.
XX
AC   AAG24052;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 27581.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1034305-A2.
```

```
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.
PR   05-MAR-1999; 99US-0123180.
PR   09-MAR-1999; 99US-0123548.
PR   23-MAR-1999; 99US-0125788.
PR   23-MAR-1999; 99US-0125793.
PR   25-MAR-1999; 99US-0126264.
PR   29-MAR-1999; 99US-0126785.
PR   01-APR-1999; 99US-0127462.
PR   01-APR-1999; 99US-0128234.
PR   06-APR-1999; 99US-0128714.
PR   08-APR-1999; 99US-0129845.
PR   16-APR-1999; 99US-0132487.
PR   19-APR-1999; 99US-0130077.
PR   21-APR-1999; 99US-0130449.
PR   23-APR-1999; 99US-0130510.
PR   23-APR-1999; 99US-0130891.
PR   28-APR-1999; 99US-0130931.
PR   30-APR-1999; 99US-0131449.
PR   30-APR-1999; 99US-0132048.
PR   04-APR-1999; 99US-0132407.
PR   04-MAY-1999; 99US-0132484.
PR   05-MAY-1999; 99US-0132485.
PR   06-MAY-1999; 99US-0132486.
PR   06-MAY-1999; 99US-0132863.
PR   07-MAY-1999; 99US-0134256.
PR   11-MAY-1999; 99US-0134218.
PR   14-MAY-1999; 99US-0134219.
PR   14-MAY-1999; 99US-0134221.
PR   14-MAY-1999; 99US-0134370.
PR   18-MAY-1999; 99US-0134768.
PR   19-MAY-1999; 99US-0134941.
PR   20-MAY-1999; 99US-0135124.
PR   21-MAY-1999; 99US-0135353.
PR   24-MAY-1999; 99US-0135629.
PR   25-MAY-1999; 99US-0136021.
PR   27-MAY-1999; 99US-0136392.
PR   28-MAY-1999; 99US-0136782.
PR   01-JUN-1999; 99US-0137222.
PR   03-JUN-1999; 99US-0137528.
PR   04-JUN-1999; 99US-0137502.
PR   07-JUN-1999; 99US-0137724.
PR   08-JUN-1999; 99US-0138094.
PR   10-JUN-1999; 99US-0138540.
PR   10-JUN-1999; 99US-0138847.
PR   14-JUN-1999; 99US-0139119.
PR   16-JUN-1999; 99US-0139452.
PR   16-JUN-1999; 99US-0139453.
PR   17-JUN-1999; 99US-0139492.
PR   18-JUN-1999; 99US-0139454.
PR   18-JUN-1999; 99US-0139455.
PR   18-JUN-1999; 99US-0139456.
PR   18-JUN-1999; 99US-0139457.
PR   18-JUN-1999; 99US-0139458.
PR   18-JUN-1999; 99US-0139459.
PR   18-JUN-1999; 99US-0139460.
PR   18-JUN-1999; 99US-0139461.
PR   18-JUN-1999; 99US-0139462.
PR   18-JUN-1999; 99US-0139463.
PR   15-JUN-1999; 99US-0139750.
PR   18-JUN-1999; 99US-0139763.
PR   21-JUN-1999; 99US-0139817.
PR   22-JUN-1999; 99US-0139899.
PR   23-JUN-1999; 99US-0140353.
PR   23-JUN-1999; 99US-0140354.
PR   24-JUN-1999; 99US-0140695.
PR   28-JUN-1999; 99US-0140823.
PR   29-JUN-1999; 99US-0140999.
PR   30-JUN-1999; 99US-0141287.
PR   01-JUL-1999; 99US-0141842.
PR   01-JUL-1999; 99US-0142154.
```

Query Match   90.1%; Score 1905.5; DB 21; Length 419;
Best Local Similarity   90.9%; Pred. No. 4.9e-157;
Matches 381; Conservative   12; Mismatches   25; Indels   1; Gaps   1;

```
Qy    1  NALVGVELKHAAEGVPE-ANCSAKPTKQGEGSLRHYYSLNIHEHQLLLRQKTHNLNRLEAQ  59
            |||||| : .|: ||   |.| | |||||.:.||:| ..|::|.||||.|:|||||||||
Db    1  MAAVGVEARPVTAMEETCNVKGAAAKQGEGLNKYYLQHLDELQRLQRERKSYNLNRLEAQ  60

Qy   60  RNDLNSRVRMLREELQLLQEPGSYVGEVVKVMGKNKVLVKVHPEGKYVVDIDKNIDITKI  119
            |.:|||||:|||||||||||:|| | |.||| |||| | ||:||||||:|| |: |  .|
Db   61  RNELNSRVRMLREELQLLQEPGSYVGEVVKVMGKNKVLVKVHPEGKYVVDIDKSIDITKL  120

Qy  120  TPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKYPDSTYDMIGGLDQQIKEIKEVIELP  179
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db  121  TPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKYPDSTYDMIGGLDQQIKEIKEVIELP  180

Qy  180  IKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKVIGEGS  239
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  IKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKVIGEGS  240

Qy  240  RMVRELFVMARBHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASN  299
            |||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  RMVRELFVMARBHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASN  300

Qy  300  KIKVLMATNRIDILDQALLRPGRIDRKIEFPTPNEESRLDILKIHSRRMNLMRGIDLKKI  359
            ||||||||||||||||||||||||||||:|:||||||||||||||||||||||||||||:
Db  301  KIKVLMATNRIDILDQALLRPGRIDRKIEFPNPNEESRFDILKIHSRKMNLMRGIDLKKI  360

Qy  360  AEKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK  418
            ||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db  361  AEKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKDTEKNMSLRKLWK  419
```

RESULT 3
AAW97652
ID   AAW97652 standard; Protein; 422 AA.
XX
AC   AAW97652;
XX
DT   10-MAY-1999 (first entry)
XX
DE   Corn SUG1 polypeptide.
XX
KW   SUG1; transcription factor; coactivator; 26S proteasome; corn;
KW   maize.
XX
OS   Zea mays.
XX
XX
PN   WO9902689-A1.
XX
PD   21-JAN-1999.
XX
PF   07-JUL-1998;
XX
PR   11-JUL-1997; 97US-0893401.
XX
PA   (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI   Allen SM, Odell JT;
XX
DR   WPI; 1999-120890/10.
DR   N-PSDB; AAX07041.
XX
PT   Newly isolated nucleic acid fragment encoding a plant SUG1 protein
PT   homologue - useful for regulating expression of specific genes
PT   normally controlled by SUG1, and for screening for plant SUG1
PT   protein inhibitors
XX
PS   Claim 1; Page 32-33; 52pp; English.
XX

CC   This polypeptide comprises the corn homologue of mouse SUG1
CC   transcriptional coactivator. Its amino acid sequence was deduced
CC   from a corn SUG1 contig (see AAX07041). SUG1 is involved in regulation
CC   of gene expression and also functions as a regulatory component of
CC   the 26S proteasome, and hence is involved in the regulation of
CC   protein turnover. Soybean, corn, wheat and rice SUG1
CC   polynucleotides (see AAX07039-43) and polypeptides (see AAW97650-54)
CC   are provided, as well as chimeric genes in which the SUG1
CC   polynucleotide is in sense or antisense orientation, and
CC   expression results in production of altered levels of SUG1 protein
CC   in transformed host cells. SUG1 proteins can be used to modulate
CC   expression of specific genes whose promoters are normally regulated
CC   by SUG1 or targeted by transcription factors that normally interact
CC   with SUG1. Host cells can be used to screen for compounds that
CC   modulate SUG1 activity.
XX
SQ   Sequence 422 AA;

Query Match 90.0%; Score 1904; DB 20; Length 422;
Best Local Similarity 90.5%; Pred. No. 6.7e-157;
Matches 383; Conservative 15; Mismatches 19; Indels 6; Gaps 3;

Qy   1 MALVGVELKH--AAEGVPEANCSAKPTK--QGEGLRHYYSLNIHEHQLLRQKTHNLNR   55
         || |: |||:.:| || |||| :||:||:|||:||||||||||
Db   1 MATVAMDISKPTPAASG-DEAAAAKGRSGGGGEGLRQYYLQHIHDLQLQIRQKTHNLNR   59

Qy   56 LEAQRNDLNSRVRMLREELQLLQEPGSYVGEVVKVKVMGKNKVLVKVHPEGRKYVVVDIDKNID 115
         ||| ||||||||||||||||||| ||||||||||||||:||||||||||:|||||||||:||
Db   60 LEAQRNDLNSRVRMLREELQLLQEPGSYVGEVVKVKVMGKSKVLVKVHPEGKXVVVDIDKSID 119

Qy 116 ITKITPSTRVALRNDSYVLHLIVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEV 175
         ||||||||||||||||:||||||||:|:||||||||||||||||||||||||||||||||||
Db 120 ITKITPSTRVALRNDSYMLHLILPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEV 179

Qy 176 IELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKYI 235
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 180 IELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKYI 239

Qy 236 GEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEYQRTMLELLNQLDGF 295
         |||||||||||||||||||||||||||||||:|||||||||:|:||||||||||||||||
Db 240 GEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGTGNGDSEVQRTMLELLNQLDGF 299

Qy 296 EASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPTPNEESRLDILKIHSRKMNLMRGID 355
         |||||||||||||||||||||||||||||||||:|:||||||||||||||:||||||||:
Db 300 EASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPNPNEDSRFDILKIHSRKMNLMRGID 359

Qy 356 LKKIAEKMNGASGGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRK 415
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 360 LKKIAEKMNGASGGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKDTEKNMSLRK 419

Qy 416 LWK 418
         |||
Db 420 LWK 422

RESULT 4
AAG24053
ID   AAG24053 standard; Protein; 405 AA.
XX
AC   AAG24053;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 27582.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 21-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.

```
PR 25-MAY-1999; 99US-0136021.          PR 03-AUG-1999; 99US-0147038.
PR 27-MAY-1999; 99US-0136392.          PR 04-AUG-1999; 99US-0147204.
PR 28-MAY-1999; 99US-0136782.          PR 04-AUG-1999; 99US-0147302.
PR 01-JUN-1999; 99US-0137222.          PR 05-AUG-1999; 99US-0147192.
PR 01-JUN-1999; 99US-0137528.          PR 05-AUG-1999; 99US-0147260.
PR 03-JUN-1999; 99US-0137502.          PR 06-AUG-1999; 99US-0147303.
PR 04-JUN-1999; 99US-0137502.          PR 06-AUG-1999; 99US-0147416.
PR 07-JUN-1999; 99US-0137724.          PR 09-AUG-1999; 99US-0147493.
PR 08-JUN-1999; 99US-0138094.          PR 09-AUG-1999; 99US-0147935.
PR 10-JUN-1999; 99US-0138540.          PR 10-AUG-1999; 99US-0148171.
PR 14-JUN-1999; 99US-0138847.          PR 11-AUG-1999; 99US-0148319.
PR 16-JUN-1999; 99US-0139119.          PR 12-AUG-1999; 99US-0148341.
PR 16-JUN-1999; 99US-0139452.          PR 13-AUG-1999; 99US-0148565.
PR 16-JUN-1999; 99US-0139453.          PR 13-AUG-1999; 99US-0148684.
PR 17-JUN-1999; 99US-0139492.          PR 16-AUG-1999; 99US-0149368.
PR 18-JUN-1999; 99US-0139454.          PR 17-AUG-1999; 99US-0149175.
PR 18-JUN-1999; 99US-0139455.          PR 18-AUG-1999; 99US-0149426.
PR 18-JUN-1999; 99US-0139456.          PR 20-AUG-1999; 99US-0149722.
PR 18-JUN-1999; 99US-0139457.          PR 20-AUG-1999; 99US-0149723.
PR 18-JUN-1999; 99US-0139458.          PR 23-AUG-1999; 99US-0149929.
PR 18-JUN-1999; 99US-0139459.          PR 23-AUG-1999; 99US-0149902.
PR 18-JUN-1999; 99US-0139460.          PR 25-AUG-1999; 99US-0149930.
PR 18-JUN-1999; 99US-0139461.          PR 26-AUG-1999; 99US-0150566.
PR 18-JUN-1999; 99US-0139462.          PR 27-AUG-1999; 99US-0150884.
PR 18-JUN-1999; 99US-0139463.          PR 27-AUG-1999; 99US-0151066.
PR 18-JUN-1999; 99US-0139750.          PR 30-AUG-1999; 99US-0151080.
PR 21-JUN-1999; 99US-0139763.          PR 31-AUG-1999; 99US-0151303.
PR 22-JUN-1999; 99US-0139817.          PR 01-SEP-1999; 99US-0151438.
PR 23-JUN-1999; 99US-0139899.          PR 07-SEP-1999; 99US-0151930.
PR 23-JUN-1999; 99US-0140353.          PR 10-SEP-1999; 99US-0152363.
PR 24-JUN-1999; 99US-0140354.          PR 13-SEP-1999; 99US-0153070.
PR 28-JUN-1999; 99US-0140695.          PR 15-SEP-1999; 99US-0153758.
PR 29-JUN-1999; 99US-0140823.          PR 16-SEP-1999; 99US-0154018.
PR 30-JUN-1999; 99US-0140991.          PR 20-SEP-1999; 99US-0154039.
PR 01-JUL-1999; 99US-0141287.          PR 22-SEP-1999; 99US-0154779.
PR 01-JUL-1999; 99US-0141842.          PR 23-SEP-1999; 99US-0155139.
PR 02-JUL-1999; 99US-0142154.          PR 24-SEP-1999; 99US-0155486.
PR 06-JUL-1999; 99US-0142055.          PR 28-SEP-1999; 99US-0155659.
PR 08-JUL-1999; 99US-0142390.          PR 29-SEP-1999; 99US-0156458.
PR 09-JUL-1999; 99US-0142803.          PR 04-OCT-1999; 99US-0156596.
PR 12-JUL-1999; 99US-0142920.          PR 05-OCT-1999; 99US-0157117.
PR 13-JUL-1999; 99US-0142977.          PR 06-OCT-1999; 99US-0157753.
PR 14-JUL-1999; 99US-0143542.          PR 07-OCT-1999; 99US-0157865.
PR 15-JUL-1999; 99US-0143624.          PR 08-OCT-1999; 99US-0158029.
PR 16-JUL-1999; 99US-0144005.          PR 12-OCT-1999; 99US-0158232.
PR 16-JUL-1999; 99US-0144085.          PR 13-OCT-1999; 99US-0158369.
PR 19-JUL-1999; 99US-0144086.          PR 13-OCT-1999; 99US-0159293.
PR 19-JUL-1999; 99US-0144325.          PR 14-OCT-1999; 99US-0159294.
PR 19-JUL-1999; 99US-0144331.          PR 14-OCT-1999; 99US-0159295.
PR 19-JUL-1999; 99US-0144332.          PR 14-OCT-1999; 99US-0159329.
PR 19-JUL-1999; 99US-0144333.          PR 14-OCT-1999; 99US-0159330.
PR 19-JUL-1999; 99US-0144334.          PR 18-OCT-1999; 99US-0159331.
PR 19-JUL-1999; 99US-0144335.          PR 21-OCT-1999; 99US-0159637.
PR 20-JUL-1999; 99US-0144352.          PR 21-OCT-1999; 99US-0159638.
PR 20-JUL-1999; 99US-0144632.          PR 21-OCT-1999; 99US-0159584.
PR 20-JUL-1999; 99US-0144884.          PR 21-OCT-1999; 99US-0160741.
PR 21-JUL-1999; 99US-0144814.          PR 21-OCT-1999; 99US-0160767.
PR 21-JUL-1999; 99US-0145086.          PR 22-OCT-1999; 99US-0160768.
PR 21-JUL-1999; 99US-0145088.          PR 22-OCT-1999; 99US-0160770.
PR 22-JUL-1999; 99US-0145087.          PR 22-OCT-1999; 99US-0160814.
PR 22-JUL-1999; 99US-0145089.          PR 25-OCT-1999; 99US-0160815.
PR 22-JUL-1999; 99US-0145192.          PR 25-OCT-1999; 99US-0160980.
PR 23-JUL-1999; 99US-0145145.          PR 25-OCT-1999; 99US-0160981.
PR 23-JUL-1999; 99US-0145218.          PR 26-OCT-1999; 99US-0160989.
PR 26-JUL-1999; 99US-0145224.          PR 26-OCT-1999; 99US-0161404.
PR 26-JUL-1999; 99US-0145276.          PR 26-OCT-1999; 99US-0161405.
PR 27-JUL-1999; 99US-0145913.          PR 28-OCT-1999; 99US-0161406.
PR 27-JUL-1999; 99US-0145918.          PR 28-OCT-1999; 99US-0161359.
PR 27-JUL-1999; 99US-0145919.          PR 28-OCT-1999; 99US-0161360.
PR 28-JUL-1999; 99US-0145951.          PR 28-OCT-1999; 99US-0161361.
PR 02-AUG-1999; 99US-0146386.          PR 28-OCT-1999; 99US-0161920.
PR 02-AUG-1999; 99US-0146388.          PR 28-OCT-1999; 99US-0161992.
PR 02-AUG-1999; 99US-0146389.
```

RESULT 5
AAW97651
ID   AAW97651 standard; Protein; 382 AA.
XX
AC   AAW97651;
XX
DT   10-MAY-1999 (first entry)
XX
DE   Corn SUG1 polypeptide.
XX
KW   SUG1; transcription factor; coactivator; 26S proteasome; corn;
KW   maize.
XX
OS   Zea mays.
XX
XX
PN   WO9902689-A1.
XX
PD   21-JAN-1999.
XX
PF   07-JUL-1998; 98WO-US13992.
XX
PR   11-JUL-1997; 97US-0893401.
XX
PA   (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI   Allen SM, Odell JT;
XX
DR   WPI; 1999-120890/10.
DR   N-PSDB; AAX07040.
XX
PT   Newly isolated nucleic acid fragment encoding a plant SUG1 protein
PT   homologue - useful for regulating expression of specific genes
PT   normally controlled by SUG1, and for screening for plant SUG1
PT   protein inhibitors
XX
PS   Claim 1; Page 28-30; 52pp; English.
XX

CC   This polypeptide comprises the corn homologue of mouse SUG1
CC   transcriptional coactivator. Its amino acid sequence was deduced
CC   from cDNA clone cs1.pk0051.b7 (see Aax07040), which was isolated
CC   on the basis of homology to mouse, yeast and Xenopus SUG1 ESTs.
CC   SUG1 is involved in regulation of gene expression and also functions
CC   as a regulatory component of the 26S proteasome, and hence is
CC   involved in the regulation of protein turnover. Soybean, corn,
CC   wheat and rice SUG1 polynucleotides (see Aax07039-43) and polypeptides
CC   (see Aaw97650-54) are provided, as well as chimeric genes in which the
CC   SUG1 polynucleotide is in sense or antisense orientation, and
CC   expression results in production of altered levels of SUG1 protein
CC   in transformed host cells. SUG1 proteins can be used to modulate
CC   expression of specific genes whose promoters are normally regulated
CC   by SUG1 or targeted by transcription factors that normally interact
CC   with SUG1. Host cells can be used to screen for compounds that
CC   modulate SUG1 activity.
XX
SQ   Sequence   382 AA;

Query Match   89.4%; Score 1891; DB 21; Length 405;
Best Local Similarity   93.1%; Pred. No. 8.5e-156;
Matches 376; Conservative 11; Mismatches 15; Indels   2; Gaps   1;

Qy   17 EANCSAK--PTKQGEGLRHYYSLNIHEHQLLLROKTHNLNRLEAQRNDLNSRVRMLREEL 74
              ||||||| |||||| |:|::||||||||||| ||||||||||||||||||||||||||
Db   2 EETCNVKGRAAAKQGEGLNKYYLQHLDELQRLQREKSYNLNRLEAQRNELNSRVRMLREEL 61

Qy   75 QLLQEPGSYVGEVVKVKVMGKNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVL 134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 QLLQEPGSYVGEVVKVKVMGKNKVLVLVKVHPEGKYVVDIDKSIDITKLTPSTRVALRNDSVL 121

Qy   135 HLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQP 194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 HLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQP 181

Qy   195 KGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKYIGEGSRMVRELFVMAREHAP 254
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   182 KGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKYIGEGSRMVRELFVMAREHAP 241

Qy   255 SIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILD 314
              |||||||||:||||||||||||||||||||||||||||||||||||||||||||||:|||
Db   242 SIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILD 301

Qy   315 QALLRPGRIDRKIEFPTPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAV 374
              |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db   302 QALLRPGRIDRKIEFPNPNEESRFDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAV 361

Qy   375 CTEAGMFALRERRVHTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
              ||||||||||||||||||||||||||||||||||||||||||
Db   362 CTEAGMFALRERRVHTQEDFEMAVAKVMKKDTEKNMSLRKLWK 405

RESULT 6
AAG24054
ID   AAG24054 standard; Protein; 350 AA.
XX
AC   AAG24054;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 27583.
XX
KW   Protein identification; signal transduction pathway; metabolic
KW   pathway; hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.

Query Match   87.9%; Score 1860; DB 20; Length 382;
Best Local Similarity   96.1%; Pred. No. 3.9e-153;
Matches 366; Conservative 11; Mismatches   4; Indels   0; Gaps   0;

Qy   38 NIHEHQLLLROKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEVVKVKVMGKNKVL 97
              :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db   2 HIHDLQQIROKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEVVKVKVMGKSKVL 61

Qy   98 VKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDS 157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 VKVHPEGKYVVDIDKSIDITKLTPSTRVALRNDSYMLHLILPSKVDPLVNLMKVEKVPDS 121

Qy   158 TYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHH 217
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 TYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHH 181

Qy   218 TDCTFIRVSGSSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNG 277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   182 TDCTFIRVSGSSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGTGNG 241

Qy   278 DSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPTPNEESR 337
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   242 DSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPNPNEDSR 301

Qy   338 LDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHTQEDFEM 397
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   302 FDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHTQEDFEM 361

Qy   398 AVAKVMKKETEKNMSLRKLWK 418
              |||||||||||||||||||||
Db   362 AVAKVMKKDTEKNMSLRKLWK 382

XX	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	18-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	21-JUN-1999;	99US-0139763.
PR	22-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140354.
PR	23-JUN-1999;	99US-0140695.
PR	24-JUN-1999;	99US-0140823.
PR	28-JUN-1999;	99US-0140991.
PR	29-JUN-1999;	99US-0141287.
PR	30-JUN-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	01-JUL-1999;	99US-0142055.
PR	02-JUL-1999;	99US-0142390.
PR	06-JUL-1999;	99US-0142803.
PR	08-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145145.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145276.
PR	26-JUL-1999;	99US-0145224.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148684.
PR	13-AUG-1999;	99US-0148865.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	25-AUG-1999;	99US-0149930.
PR	26-AUG-1999;	99US-0150566.
PR	27-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	30-AUG-1999;	99US-0151066.
PR	31-AUG-1999;	99US-0151080.
PR	01-SEP-1999;	99US-0151303.
PR	07-SEP-1999;	99US-0151438.
PR	10-SEP-1999;	99US-0151930.
PR	13-SEP-1999;	99US-0152363.
PR	15-SEP-1999;	99US-0153070.
PR	16-SEP-1999;	99US-0153758.
PR	20-SEP-1999;	99US-0154018.
PR	20-SEP-1999;	99US-0154039.
PR	22-SEP-1999;	99US-0154709.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.

PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 08-OCT-1999; 99US-0158369.
PR 12-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 82.2%; Score 1740; DB 21; Length 350;
Best Local Similarity 98.3%; Pred. No. 9.3e-143;
Matches 344; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 69 MLREELQLLQEPGSYVGEVKVKVMGKNKVLVKVHPEGKYVVDIDKNIDITKIITPSTRVALR 128
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 MLREELQLLQEPGSYVGEVVKVKVMGKNKVLVKVHPEGKYVVDIDKSIDITKLFPSTRVALR 60

Qy 129 NDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFES 188
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 NDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFES 120

Qy 189 LGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVM 248
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 LGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVM 180

Qy 249 AREHAPSIIFMDEIDSIGSARMESGSGNDSEVQRTMLELLNQLDGFEASNKIKVLMATN 308
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 AREHAPSIIFMDEIDSIGSARMESGSGNDSEVQRTMLELLNQLDGFEASNKIKVLMATN 240

Qy 309 RIDILDQALLRPGRIDRKIEFPPTPNEESSRLDILKIHSRRMNLMRGIDLKKIAEKMNGASG 368
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 RIDILDQALLRPGRIDRKIEFPNPNEESSRPDILKIHSRRMNLMRGIDLKKIAEKMNGASG 300

Qy 369 AELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 AELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKDTEKNMSLRKLWK 350

RESULT 7
ID AAB43539 standard; Protein; 402 AA.
XX
AC AAB43539;
XX
DT 08-FEB-2001 (first entry)
XX

DE Human cancer associated protein sequence SEQ ID NO:984.
XX
KW Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening.
XX
OS Homo sapiens.
XX
PN WO200055350-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05882.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587533/55.
DR N-PSDB; AAC77748.
XX
PT Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -
XX
PS Claim 11; Page 1558-1559; 2352pp; English.
XX
CC AAC77607 to AAC78448 encode the human cancer associated proteins given
CC in AAB43398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnerary; immunomodulator;
CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haematopoietic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and AAB44240 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 402 AA;

Query Match 78.9%; Score 1670.5; DB 21; Length 402;
Best Local Similarity 82.4%; Pred. No. 1.2e-136;
Matches 324; Conservative 37; Mismatches 31; Indels 1; Gaps 1;

Qy 26 KQGEGLRHYYSLNIHEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVG 85
        ||| |||:|||||:||||||:||||||:|||| |||:||:||||||||||||||||||||
Db 11 KAGSGLRQYYLSKIEELQLIVNDKSQNLRLQAQRNELNAKVRLLREELQLLQEQGGSYVG 70

Qy 86 EVVKVNGKNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPL 145
        |||| | |||:|||||||||||||||||:|||||:|| ||||||||||||||||||||
Db 71 EVVRAMDKKKVLVKVHPEGKFVVDVDKNIDLNDVTPNCRVALRNDSYTLHKILPNKVDPL 130

Qy 146 VNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGT 205
         :|| ||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db ...

```
Db  131 VSLMMVEKVPDSTYEMIGGLDKQIKEIKEVIELPVKHPELFEALGIAQPKGVLLYGPPGT 190
Qy  206 GKTLLARAVAHHTDCTFIRVSGGSELVQKYIGEGSSRMVRELFVWAREHAPSIIFMDEIDSI
                                                                      |
Qy  206 GKTLLARAVAHHTDCTFIRVSGGSELVQKYIGEGSSRMVRELFVWAREHAPSIIFMDEIDSI 265
        |||||||||||||||||||||||||||| :|||||.|||||||||||||||||||||||||
Db  191 GKTLLARAVAHHTDCTFIRVSGGSELVQKRFIGEGARMVRELFVWAREHAPSIIFMDEIDSI 250
Qy  266 GSARMESGGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDR 325
        |||  |.|.| ||||||||||||||||||||||:.||||:|||||||||||:|||||||||
Db  251 GSSRLEGGSG-GDSEVQRTMLELLNQLDGFEATKNIKVIMATNRIDILDSALLRPGRIDR 309
Qy  326 KIEFPPTPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRE 385
        ||||||.||| |||||||||||:|||.||| |.||||.|.|||||.|.||||||||.|||
Db  310 KIEFPPNEEARLDILKIHSRKMNLTRGINLRKIAELMPGASGAEVKGVCTEAGMYALRE 369
Qy  386 RRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
        |||||||||||||||||| :|.:||||:||||
Db  370 RRVHVTQEDFEMAVAKVMQKDSEKNMSIKKLWK 402
```

```
RESULT 8

AC   ABP41940;
XX
ID   ABP41940 standard; Protein; 402 AA.
XX
DT   22-AUG-2002  (first entry)
XX
DE   Human ovarian antigen HSP6G928, SEQ ID NO:3072.
XX
KW   Human; ovarian antigen; ovary; breast; cancer; tumour;
KW   ovarian cancer; breast cancer; reproductive system disorder;
KW   infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW   PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW   inflammatory condition; immune disorder; blood disorder;
KW   cardiovascular disorder; respiratory disorder; neurological disorder;
KW   gastrointestinal disorder; urinary system disorder; drug screening;
KW   gene therapy; chromosome mapping; forensic analysis;
KW   antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW   antiinflammatory; gynaecological; reproductive; chromosome 17q23-25.
XX
OS   Homo sapiens.
XX
PN   WO20020200677-A1.
XX
PD   03-JAN-2002.
XX
PF   07-JUN-2001; 2001WO-US18569.
XX
PR   07-JUN-2000; 2000US-209467P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Birse CE, Rosen CA;
XX
DR   WPI; 2002-147878/19.
DR   N-PSDB; ABQ55017.
XX
PT   Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT   useful in the prevention, treatment and diagnosis of cancer (e.g.
PT   ovarian cancer); immune disorders, cardiovascular disorders and
PT   neurological diseases -
XX
PS   Claim 11; SEQ ID No 3072; 2922pp; English.
XX
CC   The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC   ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC   encompasses polypeptides 90% identical and polynucleotides 95% identical
CC   to the sequences of the invention. The invention additionally relates to
CC   recombinant vectors and host cells comprising human ovarian antigen
CC   polynucleotides, antibodies against human ovarian antigens, and the use
CC   of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC   treating, prognosing or preventing various ovary and/or breast-related
CC   disorders. Such conditions include ovarian cancer and breast cancer, and
CC   metastatic tumours of ovarian or breast origin, reproductive system
CC   disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC   polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea) endocrine
CC   disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC   shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC   vaginitis), immune disorders (e.g., congenital and acquired
CC   immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC   blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC   respiratory disorders, neurological disorders, gastrointestinal disorders
CC   and urinary system disorders. Ovarian antigen polypeptides and
CC   polynucleotides may also be used in screening for compounds which
CC   modulate ovarian antigen expression or activity. The polynucleotides may
CC   further be used for gene therapy, chromosome mapping, in the
CC   identification of individuals and in forensic analysis, and the
CC   polypeptides may be used as food additives or to prepare antibodies
CC   useful in disease diagnosis, drug targeting and phenotyping. The present
CC   sequence represents a human ovarian antigen of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence  402 AA;
XX
```

```
Query Match             78.9%;  Score 1670.5;  DB 23;  Length 402;
Best Local Similarity   82.4%;  Pred. No. 1.2e-136;
Matches  324;  Conservative  37;  Mismatches  31;  Indels  1;  Gaps  1;

Qy   26 KQGEGLRHYYSLNITHEHQLLLRQKTHNINRLEAQRNDLNSRVRMLREELQLLQEPGSYVG 85
         ||||||   ||| :| |  |:|.|  :||||||||||||||| |||||||||.||||| ||
Db   11 KAGSGLRQYYLSKIEELQLIVNDKSQNLRRLQAQRNELNAKVRLLREELQLLQEGGSYVG 70

Qy   86 EVVKVMGKNKVLVRKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPL 145
         |||:|  |:| ||.||||.||| |||||||||||||||||||||| | |:| ||||||||
Db   81 EVVRAMDKKKVLVHPEGKFVVDVDKNIDINDVTPNCRVALRNDSYTLHKLPNKVDPL 130

Qy  146 VNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGT 205
         |:|||||||||||| ||||||:||||||||||| |||||||.||||||||||||||||||
Db  131 VSLMMVEKVPDSTYEMIGGLDKQIKEIKEVIELPVKHPELFEALGIAQPKGVLLYGPPGT 190

Qy  206 GKTLLARAVAHHTDCTFIRVSGGSELVQKYIGEGSSRMVRELFVWAREHAPSIIFMDEIDSI 265
         |||||||||||||||||||||||||||||.|||||.||||||||||||||||||||||||
Db  191 GKTLLARAVAHHTDCTFIRVSGGSELVQKRFIGEGARMVRELFVWAREHAPSIIFMDEIDSI 250

Qy  266 GSARMESGGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDR 325
         |||  |.|.| ||||||||||||||||||||:.||||:|||||||||||:|||||||||
Db  251 GSSRLEGGSG-GDSEVQRTMLELLNQLDGFEATKNIKVIMATNRIDILDSALLRPGRIDR 309

Qy  326 KIEFPPTPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRE 385
         ||||||.||| |||||||||||:|||.||| |.||||.|.|||||.|.||||||||.|||
Db  310 KIEFPPNEEARLDILKIHSRKMNLTRGINLRKIAELMPGASGAEVKGVCTEAGMYALRE 369

Qy  386 RRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
         |||||||||||||||||| :|.:||||:||||
Db  370 RRVHVTQEDFEMAVAKVMQKDSEKNMSIKKLWK 402
```

```
RESULT 9

AC   AAY27553;
XX
ID   AAY27553 standard; Protein; 406 AA.
XX
DT   03-DEC-1999  (first entry)
XX
DE   TR-interacting JL1 protein sequence.
XX
KW   JL1 protein; thyroid hormone receptor; trap assay; therapeutic;
KW   nuclear hormone receptor protein; TR-interacting protein; human;
KW   thyroid disorder.
XX
OS   Homo sapiens.
XX
PN   US5962256-A.
```

Left column (RESULT 10)

```
XX
PD    05-OCT-1999.
XX
PF    06-JUN-1995;    95US-0471613.
XX
PR    04-APR-1994;    94US-0222719.
PR    30-OCT-1992;    92US-0969136.
XX
PA    (GEHO ) GEN HOSPITAL CORP.
XX
PI    Lee JW, Moore DD;
XX
DR    WPI; 1999-571268/48.
XX
PS    Claim 1; Fig 2; 68pp; English.
XX
CC    The invention provides a JL1 protein which specifically interacts with a
CC    thyroid hormone receptor in an in vivo trap assay. The JL1 protein can be
CC    recombinantly produced by standard recombinant methodology. The JL1
CC    protein is used in a trap assay for determining whether a test protein is
CC    capable of interacting with a nuclear hormone receptor protein, and may
CC    also be used as a therapeutic peptide for treating thyroid disorders.
CC    The present sequence represents the amino acid sequence of thyroid
CC    hormone receptor (TR)-interacting protein JL1.
XX
SQ    Sequence  406 AA;
XX

Query Match             78.7%;  Score 1665;  DB 20;  Length 406;
Best Local Similarity   78.5%;  Pred. No. 3.8e-136;
Matches 328;  Conservative  38;  Mismatches  40;  Indels  12;  Gaps  2;

Qy    1  MALVGVELKHAAEGVPEANCSAKPTKQGEGLRHYSLNIHEHQLLLRQKTHNLNRLEAQR   60
         |||||||||||||||||||||||||||||||:|||:|||||||||:|||||||:|||||
Db    1  MALDGPEQMELEEG------------------------KAGSGLRQYLSKIEELQLIVNDKSQNLRRLQAQR   49

Qy   61  NDLNSRVRMLREELQLLQEPGSYVGEVAVKVMGKNKVLVKVHPEGKYVVDIDKNIDITKIT  120
         |||:||||||||||||||||||||||||||||:|||:||||||||||||||||||||||:
Db   50  NELNAKVRLLREELQLLQEQGQSYVGEVVRAMDKKKVLVKVHPEGKFVVDVDKNIDNDVT   109

Qy  121  PSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPI  180
         |:||||||||||||||||||||||||||||||||||||:||||||||||||||||||:
Db  110  PNCRVALRNDSYTLHKILPNKVDPLVSLMMVEKVPDSTYEMIGGLDKQIKEIKEVIELPV  169

Qy  181  KHPELFEALGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKFIGEGAR  240
         |||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db  170  KHPELPEALGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKFIGEGAR  229

Qy  241  MVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNK  300
         |||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db  230  MVRELFVMAREHAPSIIFMDEIDSIGSSRLEGGSG--GSSEVQRQMLELLNQLDGFEATKN  288

Qy  301  IKVLMATNRIDILDQALLRPGRIDRKIEFPTPNEESRLDILKIHSRRMNLMRGIDLKKIA  360
         |||||||||||:|||:||||||||||||||||:||||||||||||||||||||||||:
Db  289  IKVIMATNRIDMLDSALLRPGRIDRKIEFPPPNEEARLDILKIHSRKMNLTRGINLRKIA  348

Qy  361  EKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK   418
         ||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db  349  ELMPGASGAEVKGVCTEAGMYALRERRVHVTQEDFEMAVAKVMQKDSEKNMSIKKLWK   406

ID    AAW92374 standard; protein; 406 AA.
XX
AC    AAW92374;
XX
DT    21-APR-1999  (first entry)
XX
DE    Human JL1 protein.
XX

RESULT 10
AAW92374
```

Right column (RESULT 11)

```
KW    Thyroid hormone receptor-interacting protein; TR-interacting protein;
KW    JL1; human; interaction trap assay; treatment; thyroid disorder;
KW    hyperthyroidism.
XX
OS    Homo sapiens.
XX
PN    US5866686-A.
XX
PD    02-FEB-1999.
XX
PF    06-JUN-1995;    95US-0470925.
XX
PR    04-APR-1994;    94US-0222719.
PR    30-OCT-1992;    92US-0969136.
PR    06-JUN-1995;    95US-0470925.
XX
PA    (GEHO ) GEN HOSPITAL CORP.
XX
PI    Lee JW, Moore DD;
XX
DR    WPI; 1999-142013/12.
XX
PT    JL1 protein - that interacts with thyroid hormone receptor
XX
PS    Claim 1; Fig 2; 67pp; English.
XX
CC    This sequence represents a human JL1 protein that interacts with thyroid
CC    hormone receptor (TR-interacting protein) in an in vivo interaction trap
CC    assay. This protein is potentially useful for the treatment of
CC    hyperthyroidism or thyroid disorders.
XX
SQ    Sequence  406 AA;
XX

Query Match             78.7%;  Score 1665;  DB 20;  Length 406;
Best Local Similarity   78.5%;  Pred. No. 3.8e-136;
Matches 328;  Conservative  38;  Mismatches  40;  Indels  12;  Gaps  2;

Qy    1  MALVGVELKHAAEGVPEANCSAKPTKQGEGLRHYSLNIHEHQLLLRQKTHNLNRLEAQR   60
         |||||||||||||||||||||||||||||||:|||:|||||||||:|||||||:|||||
Db    1  MALDGPEQMELEEG------------------------KAGSGLRQYLSKIEELQLIVNDKSQNLRRLQAQR   49

Qy   61  NDLNSRVRMLREELQLLQEPGSYVGEVAVKVMGKNKVLVKVHPEGKYVVDIDKNIDITKIT  120
         |||:||||||||||||||||||||||||||||:|||:||||||||||||||||||||||:
Db   50  NELNAKVRLLREELQLLQEQGQSYVGEVVRAMDKKKVLVKVHPEGKFVVDVDKNIDNDVT   109

Qy  121  PSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPI  180
         |:||||||||||||||||||||||||||||||||||||:||||||||||||||||||:
Db  110  PNCRVALRNDSYTLHKILPNKVDPLVSLMMVEKVPDSTYEMIGGLDKQIKEIKEVIELPV  169

Qy  181  KHPELFEALGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKFIGEGAR  240
         |||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db  170  KHPELPEALGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKFIGEGAR  229

Qy  241  MVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNK  300
         |||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db  230  MVRELFVMAREHAPSIIFMDEIDSIGSSRLEGGSG--GSSEVQRQMLELLNQLDGFEATKN  288

Qy  301  IKVLMATNRIDILDQALLRPGRIDRKIEFPTPNEESRLDILKIHSRRMNLMRGIDLKKIA  360
         |||||||||||:|||:||||||||||||||||:||||||||||||||||||||||||:
Db  289  IKVIMATNRIDMLDSALLRPGRIDRKIEFPPPNEEARLDILKIHSRKMNLTRGINLRKIA  348

Qy  361  EKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK   418
         ||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db  349  ELMPGASGAEVKGVCTEAGMYALRERRVHVTQEDFEMAVAKVMQKDSEKNMSIKKLWK   406

ID    AAW85089 standard; protein; 406 AA.
XX
AC    AAW85089;
XX
DT    09-FEB-1999  (first entry)

RESULT 11
AAW85089
```

RESULT 12
ABB58177

Left entry

```
XX
DE   Thyroid hormone receptor-interacting protein JL1.
XX
KW   Thyroid hormone receptor-interacting protein; JL1;
KW   nuclear hormone receptor protein; screen.
XX
OS   Homo sapiens.
XX
XX
PN   US5846711-A.
XX
PD   08-DEC-1998.
XX
PF   04-APR-1994;   94US-0222719.
XX
PR   04-APR-1994;   94US-0222719.
PR   30-OCT-1992;   92US-0969136.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Lee JW, Moore DD;
XX
DR   WPI; 1999-059040/05.
XX
PT   Screening assay for nuclear hormone receptor modulators - using
PT   cells containing reporter gene construct.
XX
PS   Disclosure; Fig 2; 69pp; English.
XX
CC
CC   The present sequence represents thyroid hormone receptor-interacting
CC   protein JL1. The protein was identified using the method of the
CC   invention. The method is used to determine if a test protein is capable
CC   of interacting with a nuclear hormone receptor protein in a
CC   ligand-dependent manner. The method is used especially to screen for
CC   proteins that interact with thyroid hormone receptors in a
CC   ligand-dependent or ligand-sensitive manner.
XX
SQ   Sequence   406 AA;

Query Match             78.7%;   Score 1665;   DB 20;   Length 406;
Best Local Similarity   78.5%;   Pred. No. 3.8e-136;
Matches 328;   Conservative  38;   Mismatches  40;   Indels  12;   Gaps  2;

Qy     1  MALVGVELKHAAEGVPEANCSAKPTKQGEGLRHYSLNIHEHQLLRQKTHNLNRLEAQR  60
          |||||  ||        ||  |||| ||||| ||||| |||||||| |||| ||||||
Db     1  MALDGPEMELEEG-----------KAGSGLRQYLSKIEELQLIVNDKSQNLRRLQAQR  49

Qy    61  NDLNSRVRMLREEIQLLQEFRGSYVGEVKVKVKHPEGKVYVVDIDKNIDITKIT  120
          | ||  ||||||| ||||| | ||||| |        ||  |||||  ||||
Db    50  NELNAKVRLLREELQLLQEQGSYVGEVVRAMDKKKVLVKVHPEGKFVVVDVDKNIDINDVT  109

Qy   121  PSTRVALRNDSYVLHLVLPSKVPDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPI  180
          |  ||||||||| |  ||| ||||||| || ||||||||| ||||| ||||||||||||| |
Db   110  PNCRVALRNDSYTLHKILPNKVPDPLVSLMMVEKVPDSTYEMIGGLDKQIKEIKEVIELPV  169

Qy   181  KHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKYIGEGSR  240
          |||||| |||||||||||||||||||||||||||||||||||||||||||||| ||| |
Db   170  KHPELFEALGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKFIGEGAR  229

Qy   241  MVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNK  300
          |||||||||||||||||||||||||||  | ||| |  |  |||| ||||||||||||
Db   230  MVRELFVMAREHAPSIIFMDEIDSIGSSRLEGGSG--GSSEVQRQMLELLNQLDGFEATKN  288

Qy   301  IKVLMATNRIDILDQALLRPGRIDRKIEFPTNEESSRLDILKIHSRRMNLMRGIDLKKIA  360
          |||||||||| | ||||||||||||||||| |||  ||||||||| ||| |||| | |||
Db   289  IKVIMATNRIDMLDSALLRPGRIDRKIEFPPNEEARLDILKIHSRKMNLTRGINLRKIA  348

Qy   361  EKNNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLMK  418
          | | |||||| |||||||| ||||||||||||||||||||||| |  |||| | |||
Db   349  ELMPGASGAEVKGVCTEAGMYALRERRVHVTQEDFEMAVAKVMQKDSEKNMSIKKLWK  406
```

Right entry

```
ID   ABB58177 standard; Protein; 405 AA.
XX
AC   ABB58177;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 1323.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614450.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL02280.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 1323; 21pp + Sequence Listing; English.
XX
CC
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   405 AA;

Query Match             78.5%;   Score 1660.5;   DB 22;   Length 405;
Best Local Similarity   82.4%;   Pred. No. 9.2e-136;
Matches 323;   Conservative  39;   Mismatches  29;   Indels  1;   Gaps  1;

Qy    27  QGEGLRHYSLNIHEHQLLRQKTHNLNRLEAQRNDLNSRVRMLREEIQLLQEFRGSYVGE  86
          | || | ||     |||| | | |||| |||| |  ||||||||| ||||| | |||||
Db    15  KGEGFRSYYIQKIEELQLIVAEKHQNLRRLQAQRNELNAKVRMLREELQLLQEQGSYVGE  74

Qy    87  VVKVKVMGKNKVLVKVKHPEGKVYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVPDPLV  146
          |||   | |  ||||| ||||| | ||| |||| |||| || ||||| ||| || || ||||||||
Db    75  VVKPMDKKKVLVKVHPEGKFVVVDVDKNIDINDVTPNCRVALRNESYTLHKILPNKVPDPLV  134

Qy   147  NLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQKGVLLYGPPGTG  206
          | | ||||||||| ||||| |||||||||||||| |||||| |||||||||||||||||
Db   135  SLMMVEKVPDSTYEMIGGLDKQIKEIKEVIELPVHPELFDALGIAQPKGVLLYGPPGTG  194

Qy   207  KTLLARAVAHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIG  266
          ||||||||||| ||||||||||||||| |||||||||||||||||||||||||||||||
Db   195  KTLLARAVAHHTECTFIRVSGSELVQKFIGEGSRMVRELFVMAREHAPSIIFMDEIDSIG  254

Qy   267  SARMESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRK  326
          | | |||| | ||||||||| |||||||||||  |||||||||||||||| |||||||||
Db   255  SSRIESGSG-GDSEVQRTMELLNQLDGFEATKNIKVIMATNRIDILDPALLRPGRIDRK  313
```

```
Qy   327  IEFPPTPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRER 386
          ||||| ||||||||||||||||:||||:||||||||||||||||||:||||||||||||||
Db   314  IEFPPPNEEARLDILKIHSRKMNLTRGINLRKIAELMPGASGAEVKGVCTEAGMYALRER 373

Qy   387  RVHVTQEDFEMAVAXVMKKETEKNMSLRKLWK 418
          |||||||||||||||:||:||||||||:|||||
Db   374  RVHVTQEDFEMAVAKVMQKDSEKNMSIKKLWK 405
```

RESULT 13
AAW06100
ID   AAW06100 standard; protein; 406 AA.
XX
AC   AAW06100;
XX
DT   19-FEB-1997 (first entry)
XX
DE   Human 26S proteasome P45 subunit protein.
XX
KW   Human 26S proteasome P45 subunit; primer; PCR; polymerase chain reaction;
KW   amplification; probe; hepatoma cell line HEPG2; antibody; metabolism;
KW   immunodiagnosis; viral infection.
XX
OS   Homo sapiens.
XX
PN   JP08217796-A.
XX
PD   27-AUG-1996.
XX
PR   14-FEB-1995; 95JP-0049103.
XX
PA   14-FEB-1995; 95JP-0049103.
XX
     (SUME) SUMITOMO ELECTRIC IND CO.
XX
DR   WPI; 1996-435573/44.
DR   N-PSDB; AAT42343.
XX
PT   Human proteasome P45 subunit and coresp. DNA sequence - useful in
PT   immuno:diagnosis of e.g. viral infection and metabolic disorders
XX
XX
     Claim 1; Page 2-3; 21pp; Japanese.

This is the amino acid sequence of the P45 subunit of the human 26S
proteasome complex. The protein was isolated by conventional
chromatographic methods e.g. Sephacryl S-300, DEAE factogel,
hydroxylappatite and affinity cheromatography. The purified protein
was separated by SDS-PAGE, transferred to a PVDF membrane and used for
obtaining peptide sequence. The peptide sequence was used to design
PCR primers (AAT42344-5) which amplified a 600 bp fragment of the coding
sequence. The amplified fragment was used as a probe to isolate the full
length sequence from a cDNA library derived from mRNA purified from
human hepatoma cell line HEPG2. The protein can be used to raise
antibodies for immunodiagnosis of diseases such as viral infections and
abnormal metabolism.

SQ   Sequence   406 AA;

Query Match   78.2%; Score 1655; DB 17; Length 406;
Best Local Similarity   78.2%; Pred. No.2.8e-135;
Matches 327; Conservative 37; Mismatches 42; Indels 12; Gaps 2;

Qy   1   MALVGVELKHAAEGVPEANCSAKPTKQGEGLRHYYSLNIHEHQLLLROKTHNILNRLEAQR 60
          |||| |:|||||||:||||||:||| ||||:||||||||||||||:||||:|||||||||
Db   1   MALDGPEMELEEG------------KAGSGLROYLSKIEELQLIVDNDKSONLRRLQAQR 49

Qy   61   NDLNSRVRMLREELQLLQEPGSIYGEVKVMGKNKVLVKVHPEGKYVVDIDKNIDTKIT 120
          |:||:|||:|||||||||| |||||||:| ||||||||||||:|||:|||||| :|
Db   50   NELNAKVRLLRRELQLLQEGGSIYGEVRAMDKKVLVKVHPEGKFVVDVDKNIDINDVT 109

Qy   121 PSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEVKVPDSTIYDMIGGLDDQQIKEIKEVIELPI 180
          |. ||||||||| ||::||.|||||| :|| || |||||:| ||| | |:|||||||||||:
Db   110 PNCRVALRNDSYTLHKILPNKVDPLVSLMMVEKVPDSTYEMIGDWTKQIKEIKEVIELPV 169
```

```
Qy   181  KHPELFEESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKYIGEGSR 240
          ||||| | :||||||||||||||||||||||||||||:||:||||||||| ||:|||||| :
Db   170  KHPELPEALGIAQPKGVLLYGPPGTGKTLLARAVAVAHHTDCTFIRVSGSELVQKFIGEGAR 229

Qy   241  MVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEAASNK 300
          |||||||||||||||||||||||||||:|:|  | |  |||::|||||||||||||| : |
Db   230  MVRELFVMAREHAPSIIFMDEIDSIGSSRLEGGSG-GDSSVQRTMLELLNQLDGFEATKN 288

Qy   301  IKVLMATNRIDILDQALLRPGRIDRKIEFPPTPNEESRLDILKIHSRRMNLMRGIDLKKIA 360
          |||:||||||||||:||||||||||||||||| |||| :||||||||||||:|||:||||||
Db   289  IKVIMATNRIDILDSALLRPGRIDRKIEFPPPNEEARLDILKIHSRKMNLTRGINLRKIA 348

Qy   361  EKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
          | |.|||||| | |||||||:|||||||||||||||||||||:||.|||||||:||||
Db   349  ELMPGASGAEVKGVCTEAGMYALRERRVHVTQEDFEMAVAKVMQKDSEKNMSIKKLWK 406
```

RESULT 14
AAR53528
ID   AAR53528 standard; protein; 406 AA.
XX
AC   AAR53528;
XX
DT   25-MAR-2003 (updated)
DT   01-FEB-1995 (first entry)
XX
DE   Thyroid hormone receptor-interacting protein - JL1.
XX
KW   nuclear thyroid hormone interacting proteins; TR; JL1; JL2; SUG1;
KW   S. cerevisiae; HIV/TAT interacting proteins; MSS1; TBP1;
KW   transcriptional coactivator; treatment; diagnosis;
KW   thyroid related disorders; modulation; thyroid hormone receptor;
KW   nuclear hormone receptor; isolation.
XX
OS   Chimeric Homo sapiens.
OS   Chimeric Bacterial sp.
XX
FH   Key   Location/Qualifiers
FT   Peptide   1..49
FT   /label= N-terminal_peptide
FT   Region   45..66
FT   /label= putative leucine zipper
FT   /note= "extended by 1 hepted toward the N-terminus
FT   in this full length sequence; appears to be
FT   unique to JL1 (and SUG1)"
FT   Binding-site   190..197
FT   /label= potential ATP binding site
FT   /note= "conserved in all members of the family"
XX
PN   WO9410338-A1.
XX
PD   11-MAY-1994.
XX
PR   29-OCT-1993; 93WO-US10443.
XX
PF   30-OCT-1992; 92US-0969136.
XX
PA   (GEHO) GEN HOSPITAL CORP.
XX
PI   Lee JW, Moore DD;
XX
DR   WPI; 1994-199808/24.
XX
PT   Nuclear hormone receptor interacting polypeptides, esp. thyroid
PT   hormone=interacting proteins (TRS) - for identifying proteins
PT   useful in treatment and diagnosis of thyroid related disorders by
PT   inoculating thyroid hormone receptor activity
XX
PS   Claim 9; Page 38-39; 105pp; English.
XX
CC   This sequence shows the complete amino acid sequence of JL1, a
CC   thyroid hormone (TR) interacting protein, that has high homology
```

```
CC   to SUG1, a S.cerevisiae transcriptional coactivator and HIV/TAT
CC   interacting proteins MSS1. TR-interacting proteins physically
CC   associate with thyroid hormone receptor. Nearly all the fusion
CC   cDNAs showed very strong dependence on hormone activation. The
CC   proteins can be used in an in vivo trap system for the isolation
CC   of proteins which associate with any nuclear hormone receptor. The
CC   proteins and Abs may be used to treat or diagnose thyroid disorders,
CC   and to modulate thyroid hormone receptor activity.(See also AAR53529-57
CC   and AAQ63681-705.)
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 406 AA;

Query Match             78.2%;  Score 1654;  DB 15;  Length 406;
Best Local Similarity   78.2%;  Pred. No. 3.4e-135;
Matches 327;  Conservative 38;  Mismatches 41;  Indels 12;  Gaps 2;

Qy    1 MALVGVELKHAABGVPEANCSAKPTKQGEGLRHYYSLNIHEHQLLLRQKTHNLNRLEAQR 60
Db    1 MALDGPEQMELEGG------------------KAGSGLRQYYLSKIEELQLIVNDKSQNLRLQAQR 49

Qy   61 NDLNSRVRMLRREELQLLQEPGSYVGEVVKVWGKNKVLVKHPEGKYVVDIDKNIDITKIT 120
Db   50 NELNAKVRLLRREELQLLQEQGSYVGEVVRAMDKKKVLVKVHPEGKYVVDVDKNIDINDVT 109

Qy  121 PSTRVALRNDSTVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPI 180
Db  110 PNCRVALRNDSTTLHKILPNKVDPLVSLMMVEKVPDSTYEMIGGLDKQIKEIKEVIELPV 169

Qy  181 KHPELFEALGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKIGEGSR 240
Db  170 KHPELFEALGIAQPKGVLLYGPPGTGKTLLARAVAAHTDCTFIRVSGSELVQKFIGEGAR 229

Qy  241 MVRELFVMAREHAPSIIFMDEIDSIGSGARMESGSGNGDSEVQRTMLELLNQLDGFEASNK 300
Db  230 MVRELFVMAREHAPSIIFMDEIDSIGSSRLEGGSG-GSSEVQRQMLELLNQLDGFEATKN 288

Qy  301 IKVLMATNRIDILDQALLRPGRIDRKIEFPPTPNEESSRLDILKIHSRRMNLMRGIDLKKIA 360
Db  289 IKVIMATNRIDMLDSALLRPGRIDRKIEFPPPNEEARLDILKIHSRKMNLTRGINLRKIA 348

Qy  361 EKMNGASGGAELKAVCTEAGMFALRERRVHVTQEDEMAVARKVMKKETEKNMSLRKLWK 418
Db  349 ELMPGASGAEVKGVXTEAGMYALRERRVHVTQEDFEMAVAKVMQKDSEKNMSIKKLWK 406


RESULT 15
ID   ABB58875 standard; protein; 399 AA.
XX
AC   ABB58875;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 3417.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
```

```
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL02978.
XX
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 3417; 21pp + Sequence listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL6176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 399 AA;

Query Match             76.7%;  Score 1622.5;  DB 22;  Length 399;
Best Local Similarity   77.8%;  Pred. No. 1.8e-132;
Matches 315;  Conservative 45;  Mismatches 34;  Indels 11;  Gaps 2;

Qy   14 GVPEANCSAKPTKQGEGLRHYYSLNIHEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLRRE 73
Db    6 GIPEGS-------EGFHSYYTQKISELQFTVNERQKNJLRLQAQRNELNLKVRLREE 56

Qy   74 LQLLQEPGSYVGEVVKVWGKNKVLVKHPEGKYVVDIDKNIDITKIIPSTRVALRNDSYV 133
Db   57 LQLLQEQGSYIAEVVKPMDKNKVLVKVHPEGKYVVDVDKTINIKDVFPSSRVALRNESYT 116

Qy  134 LHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQ 193
Db  117 LHKILPNKVDPLVSLMLVEKVPDSTYEMVGGLDKQIQEIKEVIELPVKHPELFDALGITQ 176

Qy  194 PKGVLLYGPPGTGKTLLARAVAAHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHA 253
Db  177 PKGVLLYGPPGTGKTLLARAVAAHHTECTFIRVSGSELVQKFIGEGSRNVRELFVMAREHA 236

Qy  254 PSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDIL 313
Db  237 PSIIFMDEIDSIGSARLEYGT--GDSEVQRTMLELLNQLDGFEATKNIKVIMATNRIDVL 294

Qy  314 DQALLRPGRIDRKIEFPPTPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGGAELKA 373
Db  295 DQALLRPGRIDRKIEFPPPNEEARLDILKIHSRKMNLTRGINLRKIAEEMPGASGAEVKG 354

Qy  374 VCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
Db  355 VCTEAGMYALRERRVHVTQEDFEMAVSKVMMKDSEKNMSIRKFWK 399

Search completed: November  9, 2003, 12:45:37
Job time : 82 secs
```

OM protein - protein search, using sw model

Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.6

Run on: November 9, 2003, 12:44:17 ; Search time 28 seconds
(without alignments)
631.641 Million cell updates/sec

Database : Issued Patents.AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Title: US-09-462-972-2
Perfect score: 2116
Sequence: 1 MALVSVELKHAAEGVPEANC..........VAKVMKKETEKNMSLRKLMK 418

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1665	78.7	406	2	US-08-222-719-1	Sequence 1, Appli
2	1665	78.7	406	2	US-09-291-170A-7	Sequence 7, Appli
3	1665	78.7	406	2	US-09-724-884-7	Sequence 7, Appli
4	1665	78.7	406	5	PCT-US93-10443-1	Sequence 1, Appli
5	1497.5	70.8	405	2	US-08-222-719-2	Sequence 2, Appli
6	1497.5	70.8	405	2	US-08-470-925-2	Sequence 2, Appli
7	1497.5	70.8	405	2	US-08-471-613-2	Sequence 2, Appli
8	1497.5	70.8	405	5	PCT-US93-10443-2	Sequence 2, Appli
9	940.5	44.4	215	4	US-09-291-170A-6	Sequence 6, Appli
10	940.5	44.4	215	4	US-09-724-884-6	Sequence 6, Appli
11	877	41.4	389	2	US-08-820-170A-13	Sequence 13, Appl
12	877	41.4	389	3	US-09-055-699-13	Sequence 13, Appl
13	877	41.4	389	3	US-09-273-565-13	Sequence 13, Appl
14	877	41.4	389	3	US-09-565-538-13	Sequence 13, Appl
15	877	41.4	389	4	US-09-661-468-13	Sequence 13, Appl
16	877	41.4	389	4	US-09-976-165-13	Sequence 13, Appl
17	536	25.3	636	4	US-09-328-352-5495	Sequence 5495, Ap
18	533	25.2	806	3	US-08-999-774A-6	Sequence 6, Appli
19	524	24.8	413	4	US-09-149-476-751	Sequence 751, App
20	518.5	24.5	697	4	US-09-816-177-2	Sequence 2, Appli
21	515	24.3	667	4	US-09-252-991A-27180	Sequence 27180, A
22	512.5	24.2	701	4	US-09-134-001C-3327	Sequence 3327, Ap
23	511	24.1	918	4	US-09-198-452A-1072	Sequence 1072, Ap
24	488	23.1	652	3	US-08-987-123-2	Sequence 2, Appli
25	480.5	22.7	683	4	US-09-816-093-2	Sequence 2, Appli
26	480	22.7	716	4	US-09-816-093-4	Sequence 4, Appli
27	467.5	22.1	516	4	US-09-291-170A-1	Sequence 1, Appli
28	467.5	22.1	516	4	US-09-724-884-1	Sequence 1, Appl
29	465	22.0	214	4	US-09-291-170A-7	Sequence 7, Appli
30	465	22.0	214	4	US-09-724-884-7	Sequence 7, Appl
31	462.5	21.9	399	4	US-09-291-170A-4	Sequence 4, Appli
32	454.5	21.5	408	4	US-09-149-476-674	Sequence 674, App
33	447	21.1	408	4	US-09-107-532A-6253	Sequence 6253, Ap
34	433	20.5	686	2	US-08-993-228-12	Sequence 12, Appli
35	414.5	19.6	220	4	US-09-790-179-4	Sequence 4, Appli
36	407.5	19.3	217	4	US-09-291-170A-4	Sequence 4, Appli
37	407.5	19.3	217	4	US-09-724-884-4	Sequence 4, Appli
38	396.5	18.7	361	4	US-09-790-179-2	Sequence 2, Appli
39	388.5	18.4	221	4	US-09-291-170A-8	Sequence 8, Appli
40	388	18.4	221	4	US-09-724-884-8	Sequence 8, Appli
41	388	18.3	227	4	US-09-291-170A-9	Sequence 9, Appli
42	388	18.3	227	4	US-09-724-884-9	Sequence 9, Appli
43	322	15.2	213	4	US-09-291-170A-5	Sequence 5, Appli
44	322	15.2	213	4	US-09-724-884-5	Sequence 5, Appli
45	253	12.0	506	4	US-09-252-991A-27436	Sequence 27436, A

ALIGNMENTS

RESULT 1
US-08-222-719-1
Sequence 1, Application US/08222719
Patent No. 5846711
GENERAL INFORMATION:
APPLICANT: David D. Moore
APPLICANT: Jae Woon Lee
TITLE OF INVENTION: NUCLEAR HORMONE RECEPTOR-
TITLE OF INVENTION: INTERACTING POLYPEPTIDES AND
TITLE OF INVENTION: RELATED MOLECULES AND METHODS
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,719
FILING DATE: 04-April-1994
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/969,136
FILING DATE: 30-October-1992
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Paul T. Clark
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/229001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 406
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear

US-08-222-719-1

Query Match 78.7%; Score 1665; DB 2; Length 406;
Best Local Similarity 78.5%; Pred. No. 5.9e-148;
Matches 328; Conservative 38; Mismatches 40; Indels 12; Gaps 2;

Qy   1 MALVGVELKHAAEGVPEANCSAKPTKQGEGLRHYYSLNIHEHQLLLRQKTHNLNRLEAQR   60
          |:|||||||||||||||||:|||||||||||||||:|||||||||:|||||||||||
Db   1 MALDSPEQMELEEG----------------KAGSGLRQYYLSKIEELQLIVNDKSQNLRRLQAQR   49

Qy   61 NDLNSRVRMLREELQLLQEPGSYVGEVVKVMGKNKVLVKVHPEGKYVVDIDKNIDITKIT   120
          |||||:|||||||||||||:||||||||||||:||:||||||||||||:||:|||||||||:|
Db   50 NELNAXVRJLREELQLLQEQGSYVGEVVRAMDKKKVLVKVHPEGKFVVDVDKNIDINDVT   109

Qy   121 PSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPI   180
          |:||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||:
Db   110 PNCRVALRNDSYTLHKILPNKVDPLVSLMMVEKVPDSTYEMIGGLDKQIKEIKEVIELPV   169

Qy   181 KHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKYIGEGSR   240
          ||||||:|||||||||||||||||||||||||||||||||||||||:|||||||||||:|
Db   170 KHPELFEALGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKFIGEGAR   229

Qy   241 MVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNK   300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||;
Db   230 MVRELFVMAREHAPSIIFMDEIDSIGSRLEGGSG--GSSEVQRQMLELLNQLDGFEATKN   288

Qy   301 IKVLMATNRIDILDQALLRPGRIDRKIEFPTNEESRLDILKIHSRMNLMRGIDLKKIA   360
          ||:||||||||||||||||||||||||||||||||||||||:|||||||:||:|||||||:
Db   289 IKVIMATNRIDMLDSALLRPGRIDRKIEPPPNEEARLDILKIHSRKNNLTRGINLRKIA   348

Qy   361 EKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLMK   418
          |||||||||||||||||||||||||||||:||:|||||||||:||||:||:|||||||
Db   349 ELMPGASGAEVKGVCTEAGMYALRERRVHVTQEDFEMAVAKVMQKDSEKNMSIKKLWK   406

; SEQUENCE CHARACTERISTICS:
;   LENGTH: 406
;   TYPE: amino acid
;   STRANDEDNESS:
;   TOPOLOGY: N/A
US-08-470-925-1

Query Match   78.7%; Score 1665; DB 2; Length 406;
Best Local Similarity 78.5%; Pred: No. 5.9e-148;
Matches 328; Conservative 38; Mismatches 40; Indels 12; Gaps 2;

Qy   1 MALVGVELKHAAEGVPEANCSAKPTKQGEGLRHYYSLNIHEHQLLLRQKTHNLNRLEAQR   60
          |:|||||||||||||||||:|||||||||||||||:|||||||||:|||||||||||
Db   1 MALDSPEQMELEEG----------------KAGSGLRQYYLSKIEELQLIVNDKSQNLRRLQAQR   49

Qy   61 NDLNSRVRMLREELQLLQEPGSYVGEVVKVMGKNKVLVKVHPEGKYVVDIDKNIDITKIT   120
          |||||:|||||||||||||:||||||||||||:||:||||||||||||:||:|||||||||:|
Db   50 NELNAXVRJLREELQLLQEQGSYVGEVVRAMDKKKVLVKVHPEGKFVVDVDKNIDINDVT   109

Qy   121 PSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPI   180
          |:||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||:
Db   110 PNCRVALRNDSYTLHKILPNKVDPLVSLMMVEKVPDSTYEMIGGLDKQIKEIKEVIELPV   169

Qy   181 KHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKYIGEGSR   240
          ||||||:|||||||||||||||||||||||||||||||||||||||:|||||||||||:|
Db   170 KHPELFEALGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKFIGEGAR   229

Qy   241 MVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNK   300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||;
Db   230 MVRELFVMAREHAPSIIFMDEIDSIGSRLEGGSG--GSSEVQRQMLELLNQLDGFEATKN   288

Qy   301 IKVLMATNRIDILDQALLRPGRIDRKIEFPTNEESRLDILKIHSRMNLMRGIDLKKIA   360
          ||:||||||||||||||||||||||||||||||||||||||:|||||||:||:|||||||:
Db   289 IKVIMATNRIDMLDSALLRPGRIDRKIEPPPNEEARLDILKIHSRKNNLTRGINLRKIA   348

Qy   361 EKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLMK   418
          |||||||||||||||||||||||||||||:||:|||||||||:||||:||:|||||||
Db   349 ELMPGASGAEVKGVCTEAGMYALRERRVHVTQEDFEMAVAKVMQKDSEKNMSIKKLWK   406

RESULT 2
US-08-470-925-1
; Sequence 1, Application US/08470925
; Patent No. 5866686
; GENERAL INFORMATION:
;   APPLICANT: David D. Moore
;   TITLE OF INVENTION: NUCLEAR HORMONE RECEPTOR-
;   TITLE OF INVENTION: INTERACTING POLYPEPTIDES AND
;   TITLE OF INVENTION: RELATED MOLECULES AND METHODS
;   NUMBER OF SEQUENCES: 31
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson P.C.
;   STREET: 225 Franklin Street
;   CITY: Boston
;   STATE: Massachusetts
;   COUNTRY: U.S.A.
;   ZIP: 02110-2804
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;   COMPUTER: IBM PS/2 Model 50z or 55SX
;   OPERATING SYSTEM: MS-DOS (Version 5.0)
;   SOFTWARE: WordPerfect (Version 5.1)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/470,925
;   FILING DATE: 06-June-1995
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/222,719
;   FILING DATE: 04-April-1994
;   CLASSIFICATION: 435
;   APPLICATION NUMBER: 07/969,136
;   FILING DATE: October 30, 1992
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Paul T. Clark
;   REGISTRATION NUMBER: 30,162
;   REFERENCE/DOCKET NUMBER: 00786/229003
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 542-5070
;   TELEFAX: (617) 542-8906
;   TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:

RESULT 3
US-08-471-613-1
; Sequence 1, Application US/08471613
; Patent No. 5962256
; GENERAL INFORMATION:
;   APPLICANT: David D. Moore
;   APPLICANT: Jae Woon Lee
;   TITLE OF INVENTION: NUCLEAR HORMONE RECEPTOR-
;   TITLE OF INVENTION: INTERACTING POLYPEPTIDES AND
;   TITLE OF INVENTION: RELATED MOLECULES AND METHODS
;   NUMBER OF SEQUENCES: 31
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson P.C.
;   STREET: 225 Franklin Street
;   CITY: Boston
;   STATE: Massachusetts
;   COUNTRY: U.S.A.
;   ZIP: 02110-2804
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;   COMPUTER: IBM PS/2 Model 50z or 55SX
;   OPERATING SYSTEM: MS-DOS (Version 5.0)
;   SOFTWARE: WordPerfect (Version 5.1)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/471,613
;   FILING DATE: 06-June-1995
;   CLASSIFICATION: 436
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/222,719
;   FILING DATE: 04-April-1994
;   CLASSIFICATION: 436
;   APPLICATION NUMBER: 07/969,136
;   FILING DATE: October 30, 1992

Left column (RESULT 4)

CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Paul T. Clark
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/229001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEX: 200154
TELEFAX: (617) 542-8906
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 406
TYPE: amino acid
STRANDEDNESS: N/A
TOPOLOGY: N/A
US-08-471-613-1

Query Match   78.7%; Score 1665; DB 2; Length 406;
Best Local Similarity   78.5%; Pred. No. 5.9e-148;
Matches 328; Conservative 38; Mismatches 40; Indels 12; Gaps 2;

```
Qy    1 MALVGVELKHAAEGVPEANCSAKPTKQEGLRHYYSLNIHEHQLLLRQKTHNLNRLEAQR 60
Db    1 MALDGPEQMELEEG---------KAGSGLRQYLSKIEELQLIVNDKSQNLRRLQAQR 49

Qy   61 NDLNSRVRMLREEELQLLQEPGSYVGEVKVKVMGKNKVLVKVHPEGKYVVDIDKNIDITKIT 120
Db   50 NELNAKVRLLREEELQLLQEQGSYVGEVVRAMDKKKVLVKVHPEGKFVVDVDKNIDINDVT 109

Qy  121 PSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQIKEIKEVIELP1 180
Db  110 PNCRVALRNDSYTLHKILPNKVDPLVSLMVVEKVPDSTYEMIGGLDRQIKEIKEVIELPV 169

Qy  181 KHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKYIGEGSR 240
Db  170 KHPELFEALGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKFIGEGAR 229

Qy  241 MVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNK 300
Db  230 MVRELFVMAREHAPSIIFMDEIDSIGSSRLEGGSG-GSSEVQRQMLELLNQLDGFEATKN 288

Qy  301 IKVLMATNRIDILDQALLRPGRIDRKIEFPTTPNEESSRLDILKIHSRRMNLMRGIDLKKIA 360
Db  289 IKVIMATNRIDMLDSALLRPGRIDRKIEPFPPPNEEARLDILKIHSRKMNLTRGINLRKIA 348

Qy  361 EKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
Db  349 ELMPGASGAEVKGVCTEAGMYALRERRVHVTQEDFEMAVAKVMQKDSEKNMSIKKLWK 406
```

RESULT 4
PCT-US93-10443-1
; Sequence 1, Application PC/TUS9310443
; GENERAL INFORMATION:
; APPLICANT: David D. Moore
; APPLICANT: Jae W. Lee
; TITLE OF INVENTION: NUCLEAR HORMONE RECEPTOR-
; TITLE OF INVENTION: INTERACTING POLYPEPTIDES AND
; TITLE OF INVENTION: RELATED MOLECULES AND METHODS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:

Right column (RESULT 5)

APPLICATION NUMBER: PCT/US93/10443
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/969,136
FILING DATE: October 30, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Paul T. Clark
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/099002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEX: 200154
TELEFAX: (617) 542-8906
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 406
TYPE: amino acid
STRANDEDNESS: N/A
TOPOLOGY: N/A
PCT-US93-10443-1

Query Match   78.7%; Score 1665; DB 5; Length 406;
Best Local Similarity   78.5%; Pred. No. 5.9e-148;
Matches 328; Conservative 38; Mismatches 40; Indels 12; Gaps 2;

```
Qy    1 MALVGVELKHAAEGVPEANCSAKPTKQEGLRHYYSLNIHEHQLLLRQKTHNLNRLEAQR 60
Db    1 MALDGPEQMELEEG---------KAGSGLRQYLSKIEELQLIVNDKSQNLRRLQAQR 49

Qy   61 NDLNSRVRMLREEELQLLQEPGSYVGEVKVKVMGKNKVLVKVHPEGKYVVDIDKNIDITKIT 120
Db   50 NELNAKVRLLREEELQLLQEQGSYVGEVVRAMDKKKVLVKVHPEGKFVVDVDKNIDINDVT 109

Qy  121 PSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQIKEIKEVIELP1 180
Db  110 PNCRVALRNDSYTLHKILPNKVDPLVSLMVVEKVPDSTYEMIGGLDRQIKEIKEVIELPV 169

Qy  181 KHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKYIGEGSR 240
Db  170 KHPELFEALGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKFIGEGAR 229

Qy  241 MVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNK 300
Db  230 MVRELFVMAREHAPSIIFMDEIDSIGSSRLEGGSG-GSSEVQRQMLELLNQLDGFEATKN 288

Qy  301 IKVLMATNRIDILDQALLRPGRIDRKIEFPTTPNEESSRLDILKIHSRRMNLMRGIDLKKIA 360
Db  289 IKVIMATNRIDMLDSALLRPGRIDRKIEPFPPPNEEARLDILKIHSRKMNLTRGINLRKIA 348

Qy  361 EKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
Db  349 ELMPGASGAEVKGVCTEAGMYALRERRVHVTQEDFEMAVAKVMQKDSEKNMSIKKLWK 406
```

RESULT 5
US-08-222-719-2
; Sequence 2, Application US/08222719
; Patent No. 5846712
; GENERAL INFORMATION:
; APPLICANT: David D. Moore
; APPLICANT: Jae Woon Lee
; TITLE OF INVENTION: NUCLEAR HORMONE RECEPTOR-
; TITLE OF INVENTION: INTERACTING POLYPEPTIDES AND
; TITLE OF INVENTION: RELATED MOLECULES AND METHODS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804

; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/222,719
; FILING DATE: 04-April-1994
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/969,136
; FILING DATE: 30-October-1992
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/229001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 405
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-222-719-2

Query Match   70.8%; Score 1497.5; DB 2; Length 405;
Best Local Similarity   74.0%; Pred. No. 3.1e-132;
Matches 288; Conservative 49; Mismatches 51; Indels 1; Gaps 1;

Qy   30 GLRHYSLNIHEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSSYVGEVVK 89
         ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||:|
Db   18 GIKPYPEQKIQETELKIRSKTENGRRLEAQRNALNDKVRFIKDELRLLQEPGSSYVGEVIK 77

Qy   90 VMGKNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLM 149
         :: |||||||||||||||||||||||||:|||||:||:|
Db   78 IVSDKKVLVKVQPEGKYIVDVAKDINVKDLKASQRVCLRSDSYMLHKVLENKADPLVSIM 137

Qy   150 KVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTL 209
         |:|:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   138 MVEKVPDSTYDMVGGLTKQIKEIKEVIELPVKHPELFESLGIAQPKGVILYGPPGTGKTL 197

Qy   210 LARAVAHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVNAREHAPSIIFMDEIDSIGSAR 269
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   198 LARAVAHHTDCKFIRVSGAELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSTR 257

Qy   270 MESGSGNGDGSEVQRTMLELLNQLDGFEAASNKIKVLMATNRIDILDQALLRPGRIDRKIEF 329
         ||:|||||||:||:|||||||||||||:||||||||||||||||||||||||||||||||:|
Db   258 VE-GSGGDSEVQRTMLELLNQLDGFETSKNIKIIMATNRLDILDPALLRPGRIDRKIEF 316

Qy   330 PTPNEESRLDILKIHSRRMNLMRGIDLKKIIAEKMNGASGAELKAVCTEAGMFALRERRVH 389
         |||||||||||||||||||||||||||:|||||||||:||::|||||||||||||:|:||:|
Db   317 PPPSVAARAELIRIHSRKMNLTRGINLRKVAEKNNGCSGADVKGVCTEAGMYALRERRIH 376

Qy   390 VTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
         |||||||||||||||||||||||||||||
Db   377 VTQEDFELAVGKVMNKNQETAISVAKLFK 405

; RESULT 6
; US-08-470-925-2
; Sequence 2, Application US/08470925
; Patent No. 5866686
; GENERAL INFORMATION:
; APPLICANT: Jae Woon Lee
; TITLE OF INVENTION: NUCLEAR HORMONE RECEPTOR-
; TITLE OF INVENTION: INTERACTING POLYPEPTIDES AND
; TITLE OF INVENTION: RELATED MOLECULES AND METHODS
; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,925
; FILING DATE: 06-June-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/222,719
; FILING DATE: 04-April-1994
; CLASSIFICATION: 435
; APPLICATION NUMBER: 07/969,136
; FILING DATE: October 30, 1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/229003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 405
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: N/A
; US-08-470-925-2

Query Match   70.8%; Score 1497.5; DB 2; Length 405;
Best Local Similarity   74.0%; Pred. No. 3.1e-132;
Matches 288; Conservative 49; Mismatches 51; Indels 1; Gaps 1;

Qy   30 GLRHYSLNIHEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSSYVGEVVK 89
         ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||:|
Db   18 GIKPYPEQKIQETELKIRSKTENGRRLEAQRNALNDKVRFIKDELRLLQEPGSSYVGEVIK 77

Qy   90 VMGKNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLM 149
         :: |||||||||||||||||||||||||:|||||:||:|
Db   78 IVSDKKVLVKVQPEGKYIVDVAKDINVKDLKASQRVCLRSDSYMLHKVLENKADPLVSIM 137

Qy   150 KVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTL 209
         |:|:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   138 MVEKVPDSTYDMVGGLTKQIKEIKEVIELPVKHPELFESLGIAQPKGVILYGPPGTGKTL 197

Qy   210 LARAVAHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVNAREHAPSIIFMDEIDSIGSAR 269
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   198 LARAVAHHTDCKFIRVSGAELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSTR 257

Qy   270 MESGSGNGDGSEVQRTMLELLNQLDGFEAASNKIKVLMATNRIDILDQALLRPGRIDRKIEF 329
         ||:|||||||:||:|||||||||||||:||||||||||||||||||||||||||||||||:|
Db   258 VE-GSGGDSEVQRTMLELLNQLDGFETSKNIKIIMATNRLDILDPALLRPGRIDRKIEF 316

Qy   330 PTPNEESRLDILKIHSRRMNLMRGIDLKKIIAEKMNGASGAELKAVCTEAGMFALRERRVH 389
         |||||||||||||||||||||||||||:|||||||||:||::|||||||||||||:|:||:|
Db   317 PPPSVAARAELIRIHSRKMNLTRGINLRKVAEKNNGCSGADVKGVCTEAGMYALRERRIH 376

Qy   390 VTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
         |||||||||||||||||||||||||||||
Db   377 VTQEDFELAVGKVMNKNQETAISVAKLFK 405

RESULT 7

US-08-471-613-2
; Sequence 2, Application US/08471613
; Patent No. 5962256
; GENERAL INFORMATION:
; APPLICANT: David D. Moore
; APPLICANT: Jae Woon Lee
; TITLE OF INVENTION: NUCLEAR HORMONE RECEPTOR-
; TITLE OF INVENTION: INTERACTING POLYPEPTIDES AND
; TITLE OF INVENTION: RELATED MOLECULES AND METHODS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/222,719
; FILING DATE: 04-April-1994
; CLASSIFICATION: 436
; APPLICATION NUMBER: 436
; FILING DATE: 07/969,136
; FILING DATE: October 30, 1992
; CLASSIFICATION: 436
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,613
; FILING DATE: 06-June-1995
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/229001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 405
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: N/A
US-08-471-613-2

```
Query Match     70.8%;  Score 1497.5; DB 2; Length 405;
Best Local Similarity   74.0%;  Pred. No. 3.1e-132;
Matches  288; Conservative  49; Mismatches  51; Indels  1; Gaps  1;

Qy    30 GLRHYYSLNIHEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSSYVGEVVK  89
         | :|  ||  |: ||||||:||  ||||||| |:  ||| |:|||||||||||  ||||: |
Db    18 GIKPYFEQKIQETELKIRSKTENGRRLEAQRNALNDKVRFIKDELRLLQEPGSYVGEVIK  77

Qy    90 VMGKNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLM  149
         :|  :|||||| |||||:||: |:|:| | |  :|| ||||||:||:||  | |||||::|
Db    78 IVSDKKVLVKVQPEGKYIVDVAKDINVKDLKASQRVCLRSDSYMLHKVLENKADPLVSIM  137

Qy   150 KVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTL  209
         :||||||||||| ||| :||||||||||||:|||||||||||||||| :|||||||||||
Db   138 MVEKVPDSTYDMVGGLTKQIKEIKEVIELPVKHPELFESLGIAQPKGVILYGPPGTGKTL  197

Qy   210 LARAVAHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSAR  269
         |||||||||| ||||||| |||||||||||||||||||||||||||||||||:|||||| |
Db   198 LARAVAHHTDCKFIRVSGAELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSTR  257

Qy   270 MESGSGGNGDSEVQRTMLELLNQLDGFEASNKIVLMATNRIDILDQALLRPGRIDRKIEF  329
         :|:|||| ||||||:|||||||||||:||  |||: |||||||||: |||||||||||||
Db   258 VE-GSGGDSEVQRTMLELLNQLDGFETSKNIKIMATNRLDILDPALLRPGRIDRKIEF  316
```

Qy   330 PTPNEESRLDILKIHSRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVH 389
         : | | ||:||::|||||||||:|| :||:||||||| |:|:||||||| ||||| :|
Db   317 PPPSVAARAEILRIHSRKMNLTRGINLRKVAEKMNGCSGADVKGVCTEAGMYALRERIH 376

Qy   390 VTQEDFEMAVAKYMKSTEKNMSLRKLWK 418
         |||||||:||.|.| |.:   :|||.|
Db   377 VTQEDFELAVGKVMNKNQETAISVAKLFK 405

RESULT 8
PCT-US93-10443-2
; Sequence 2, Application PCT/US9310443
; GENERAL INFORMATION:
; APPLICANT: David D. Moore
; APPLICANT: Jae W. Lee
; TITLE OF INVENTION: NUCLEAR HORMONE RECEPTOR-
; TITLE OF INVENTION: INTERACTING POLYPEPTIDES AND
; TITLE OF INVENTION: RELATED MOLECULES AND METHODS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/10443
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/969,136
; FILING DATE: October 30, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/099002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 405
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
PCT-US93-10443-2

```
Query Match     70.8%;  Score 1497.5; DB 5; Length 405;
Best Local Similarity   74.0%;  Pred. No. 3.1e-132;
Matches  288; Conservative  49; Mismatches  51; Indels  1; Gaps  1;

Qy    30 GLRHYYSLNIHEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSSYVGEVVK  89
         | :|  ||  |: ||||||:||  ||||||| |:  ||| |:|||||||||||  ||||: |
Db    18 GIKPYFEQKIQETELKIRSKTENGRRLEAQRNALNDKVRFIKDELRLLQEPGSYVGEVIK  77

Qy    90 VMGKNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLM  149
         :|  :|||||| |||||:||: |:|:| | |  :|| ||||||:||:||  | |||||::|
Db    78 IVSDKKVLVKVQPEGKYIVDVAKDINVKDLKASQRVCLRSDSYMLHKVLENKADPLVSIM  137

Qy   150 KVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTL  209
         :||||||||||| ||| :||||||||||||:|||||||||||||||| :|||||||||||
Db   138 MVEKVPDSTYDMVGGLTKQIKEIKEVIELPVKHPELFESLGIAQPKGVILYGPPGTGKTL  197

Qy   210 LARAVAHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSAR  269
         |||||||||| ||||||| |||||||||||||||||||||||||||||||||:|||||| |
Db   198 LARAVAHHTDCKFIRVSGAELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSTR  257
```

```
Qy  270 MESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRXIEF 329
        |:|||||||||| ||||||||||||||||||||||||||||||:||||||||||||||||
Db  258 VE-GSGGGDSEVQRTMLELLNQLDGFETSKNIKIIMATNRLDILDPALLRPGRIDRKIEF 316

Qy  330 PTPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVH 389
        || :||||:|||||||||||||||||:| |||||||||:|||| ||||:|||||||||:|
Db  317 PPSVAARAEILRIHSRKMNLTRGINLRKVAEKMNGCSGADVKGVCTEAGMYALRERRIH 376

Qy  390 VTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
        ||||||| ||.||||:|||.|:||:||:|
Db  377 VTQEDFELAVGKVNMKNQETAISVAKLFK 405
```

RESULT 9
US-09-291-170A-6
Sequence 6, Application US/09291170A
Patent No. 6410687
GENERAL INFORMATION:
APPLICANT: Vale, Ronald D.
APPLICANT: Hartman, James J.
APPLICANT: The Regents of the University of California
TITLE OF INVENTION: Assays for the Detection of Microtubule
TITLE OF INVENTION: Depolymerization Inhibitors
FILE REFERENCE: 18557B-0005100S
CURRENT APPLICATION NUMBER: US/09/291,170A
CURRENT FILING DATE: 1999-04-13
PRIOR APPLICATION NUMBER: US 60/081,734
PRIOR FILING DATE: 1998-04-14
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
  LENGTH: 215
  TYPE: PRT
  ORGANISM: Saccharomyces cerevisiae
  FEATURE:
  OTHER INFORMATION: AAA ATPase superfamily sug1 AAA domain
US-09-291-170A-6

Query Match   44.4%; Score 940.5; DB 4; Length 215;
Best Local Similarity   84.3%; Pred. No. 2.2e-80;
Matches 182; Conservative 19; Mismatches 14; Indels 1; Gaps 1;

```
Qy  157 STYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAH 216
        ||||| :||| ||||||||||||:|||||||||||||||||:|||||||||||||||||
Db    1 STYDMVGGLTKQIKEIKEVIELPVKHPELFESLGIAQPKGVILYGPPGTGKTLLARAVAH 60

Qy  217 HTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGN 276
        |||| |||||| :|||||:|:||||||||||||||||||||||||||||||||: |  || |
Db   61 HTDCKFIRVSGAELVQKIIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSTRVE-GSGG 119

Qy  277 GDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPTNEES  336
        |||||||||||||||||||||:| ||||:||||:|||||:||||||||||||||||: :
Db  120 GDSEVQRTMLELLNQLDGFETSKNIKIIMATNRLDILDPALLRPGRIDRKIEFPPSVAA  179

Qy  337 RLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELK 372
        | :|| |||:|||| |:|:|||||||||| ||||:.|
Db  180 RAEILRIHSRKMNLTRGINLRKVAEKMNGCSGADVK 215
```

RESULT 10
US-09-724-884-6
Sequence 6, Application US/09724884
Patent No. 6429304
GENERAL INFORMATION:
APPLICANT: Vale, Ronald D.
APPLICANT: Hartman, James J.
APPLICANT: The Regents of the University of California
TITLE OF INVENTION: Assays for the Detection of Microtubule
TITLE OF INVENTION: Depolymerization Inhibitors
FILE REFERENCE: 18557B-0005100S
CURRENT APPLICATION NUMBER: US/09/724,884
CURRENT FILING DATE: 2000-11-28

PRIOR APPLICATION NUMBER: 09/291,170
PRIOR FILING DATE: 1999-04-13
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
  LENGTH: 215
  TYPE: PRT
  ORGANISM: Saccharomyces cerevisiae
  FEATURE:
  OTHER INFORMATION: AAA ATPase superfamily sug1 AAA domain
US-09-724-884-6

Query Match   44.4%; Score 940.5; DB 4; Length 215;
Best Local Similarity   84.3%; Pred. No. 2.2e-80;
Matches 182; Conservative 19; Mismatches 14; Indels 1; Gaps 1;

```
Qy  157 STYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAH 216
        ||||| :||| ||||||||||||:|||||||||||||||||:|||||||||||||||||
Db    1 STYDMVGGLTKQIKEIKEVIELPVKHPELFESLGIAQPKGVILYGPPGTGKTLLARAVAH 60

Qy  217 HTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGN 276
        |||| |||||| :|||||:|:||||||||||||||||||||||||||||||||: |  || |
Db   61 HTDCKFIRVSGAELVQKIIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSTRVE-GSGG 119

Qy  277 GDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPTNEES  336
        |||||||||||||||||||||:| ||||:||||:|||||:||||||||||||||||: :
Db  120 GDSEVQRTMLELLNQLDGFETSKNIKIIMATNRLDILDPALLRPGRIDRKIEFPPSVAA  179

Qy  337 RLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELK 372
        | :|| |||:|||| |:|:|||||||||| ||||:.|
Db  180 RAEILRIHSRKMNLTRGINLRKVAEKMNGCSGADVK 215
```

RESULT 11
US-08-820-170A-13
Sequence 13, Application US/08820170A
Patent No. 5831058
GENERAL INFORMATION:
APPLICANT: Tsutomu, FUJIWARA
APPLICANT: Takeshi, WATANABE
APPLICANT: Masato, HORIE
APPLICANT: Toyonasa, KATAGIRI
TITLE OF INVENTION: HUMAN GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
  ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
  STREET: 2100 Pennsylvania Avenue, N.W.
  CITY: Washington
  STATE: D.C.
  COUNTRY: United States
  ZIP: 20037-3202
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/820,170A
  FILING DATE:
  CLASSIFICATION: 536
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (202) 293-7060
  TELEFAX: (202) 293-7860
  TELEX: 6491103
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
  LENGTH: 389 amino acids
  TYPE: amino acid
  TOPOLOGY: linear
  MOLECULE TYPE: protein
US-08-820-170A-13

Query Match   41.4%; Score 877; DB 2; Length 389;

Best Local Similarity 48.0%; Pred. No. 5e-74; Length 389;
Matches 183; Conservative 64; Mismatches 128; Indels 6; Gaps 3;

```
Qy   35 YSLNIHEHQLLRQKTHNLRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEVVKVMGKN  94
Db   12 YRKLLEH---KEIDGRLKELREQLKELTKQYEKSENDLKALQSVGQIVGEVLKQLTEE  67

Qy   95 KVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKV 154
Db   68 KFIVKATNGPRYVVGCRRQLDKSKLKPGTRVALDMTLTIMRYLPREVDPLVYNMSHEDP 127

Qy  155 PDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAV 214
Db  128 GNVSYSEIGGLSEQIRELREVIELPLTNPELFQRVGIIPPKGCLLYGPPGTGKTLLARAV 187

Qy  215 AHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGS 274
Db  188 ASQLDCNFLKVVSSSIVDKYIGESARLIREMFNYARDHQPCIIFMDEIDAIGGRRFSEGT 247

Qy  275 GNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPTNE 334
Db  248 -SADREIQRTLMELLNQMDGFDTLHRVKMTMATNRPDTLDPALLRPGRLDRKIHIDLPNE 306

Qy  335 ESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGGAELKAVCTEAGMFALRERRVHVTQED 394
Db  307 QARLDILKIHAGPITKHGEIDYEAIVKLSDGFNGADLRNVCTEAGMFAIRADHDFVVQED 366

Qy  395 FEMAVAKVM-KKETEKNMSLR 414
Db  367 FMKAVRKVADSKKLESKLDYK 387
```

RESULT 12
US-09-055-699-13
Sequence 13, Application US/09055699
Patent No. 6005088
GENERAL INFORMATION:
APPLICANT: Tsutomu, FUJIWARA
APPLICANT: Takeshi, WATANABE
APPLICANT: Masato, HORIE
APPLICANT: Toyomasa, KATAGIRI
TITLE OF INVENTION: HUMAN GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/055,699
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/820,170
FILING DATE:
CLASSIFICATION:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 649103
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 389 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-055-699-13

Query Match 41.4%; Score 877; DB 3; Length 389;
Best Local Similarity 48.0%; Pred. No. 5e-74; Length 389;
Matches 183; Conservative 64; Mismatches 128; Indels 6; Gaps 3;

```
Qy   35 YSLNIHEHQLLRQKTHNLRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEVVKVMGKN  94
Db   12 YRKLLEH---KEIDGRLKELREQLKELTKQYEKSENDLKALQSVGQIVGEVLKQLTEE  67

Qy   95 KVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKV 154
Db   68 KFIVKATNGPRYVVGCRRQLDKSKLKPGTRVALDMTLTIMRYLPREVDPLVYNMSHEDP 127

Qy  155 PDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAV 214
Db  128 GNVSYSEIGGLSEQIRELREVIELPLTNPELFQRVGIIPPKGCLLYGPPGTGKTLLARAV 187

Qy  215 AHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGS 274
Db  188 ASQLDCNFLKVVSSSIVDKYIGESARLIREMFNYARDHQPCIIFMDEIDAIGGRRFSEGT 247

Qy  275 GNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPTNE 334
Db  248 -SADREIQRTLMELLNQMDGFDTLHRVKMTMATNRPDTLDPALLRPGRLDRKIHIDLPNE 306

Qy  335 ESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGGAELKAVCTEAGMFALRERRVHVTQED 394
Db  307 QARLDILKIHAGPITKHGEIDYEAIVKLSDGFNGADLRNVCTEAGMFAIRADHDFVVQED 366

Qy  395 FEMAVAKVM-KKETEKNMSLR 414
Db  367 FMKAVRKVADSKKLESKLDYK 387
```

RESULT 13
US-09-273-565-13
Sequence 13, Application US/09273565A
Patent No. 6166190
GENERAL INFORMATION:
APPLICANT: FUJIWARA, TSUTOMU
APPLICANT: WATANABE, TAKESHI
APPLICANT: HORIE, MASATO
TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
FILE REFERENCE: Q-53599
CURRENT APPLICATION NUMBER: US/09/273,565A
CURRENT FILING DATE: 1999-03-22
EARLIER APPLICATION NUMBER: 09/055,699
EARLIER FILING DATE: 1998-04-07
EARLIER APPLICATION NUMBER: 08/820,170
EARLIER FILING DATE: 1997-03-19
EARLIER APPLICATION NUMBER: JP 63410/1996
EARLIER FILING DATE: 1996-03-19
EARLIER APPLICATION NUMBER: JP 69163/1997
EARLIER FILING DATE: 1997-03-05
NUMBER OF SEQ ID NOS: 95
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 13
LENGTH: 389
TYPE: PRT
ORGANISM: Homo sapiens
US-09-273-565-13

```
RESULT 14
US-09-565-538-13
Sequence 13, Application US/0956538
; Patent No. 6333404
; GENERAL INFORMATION:
;  APPLICANT: FUJIWARA, TSUTOMU
;  APPLICANT: WATANABE, TAKESHI
;  APPLICANT: HORIE, MASATO
;  TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
;  TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
;  FILE REFERENCE: Q-53599
;  CURRENT APPLICATION NUMBER: US/09/565,538
;  CURRENT FILING DATE: 2000-05-05
;  PRIOR APPLICATION NUMBER: 09/273,565
;  PRIOR FILING DATE: 1999-03-22
;  PRIOR APPLICATION NUMBER: 09/055,699
;  PRIOR FILING DATE: 1998-04-07
;  PRIOR APPLICATION NUMBER: 08/820,170
;  PRIOR FILING DATE: 1997-03-19
;  PRIOR APPLICATION NUMBER: JP 63410/1996
;  PRIOR FILING DATE: 1996-03-19
;  PRIOR APPLICATION NUMBER: JP 69163/1997
;  PRIOR FILING DATE: 1997-03-05
;  NUMBER OF SEQ ID NOS: 95
;  SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 13
;  LENGTH: 389
;  TYPE: PRT
;  ORGANISM: Homo sapiens
US-09-565-538-13

Query Match             41.4%;  Score 877;  DB 4;  Length 389;
Best Local Similarity   48.0%;  Pred. No. 5e-74;
Matches  183;  Conservative   64;  Mismatches 128;  Indels    6;  Gaps    3;

Qy    35 YSLNIHEHQLLRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEVAVKVMGKN  94
            :  |||  ||         || :|  :|  :  |  ||| |||:||:|| |:  :
Db    12 YRKKLLEH---KEIDGRLKELREQLKELTKQYEKSENDLKALQSVGQIVGEVLKQLTEE   67

Qy    95 KVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKV 154
            || |   |||:|||   |  || |||||||||:    :|  |  |||||||:||    :
Db    68 KFIVKATNGPRVVGCRRQLDKSKLKPGTRVALDMTTLIMRYLPREVDPLVINMSHEDP  127

Qy   155 PDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAV 214
           :|::| :|||| :||:|:||||||| ::|||| :|||:||||||||||||||||||||
Db   128 GNVSYSEIGGLSEQIRELREVIELPLTNPELFQRVGIIPPKGCLLYGPPGTGKTLLARAV 187

Qy   215 AHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGS 274
           |  ||| |:|||::|||||||||:||:||:| : ||:|| |||||||||:||:||| ||:
Db   188 ASQLDCNFLKVVSSIVDKYIGESARLIREMFNYARDHQPCIIFMDEIDAIGGRRFSEGT  247

Qy   275 GNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPPTNE 334
             ::| :|||:|||||||:|||: : :|||:|||||: |||:|||||||:|||| | |||
Db   248 -SADREIQRTLMELLNQMDGFDTLHRVKVTMATNRPDTLDPALLRPGRLDRKIHIDLPNE 306

Qy   335 ESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHVTQED 394
           ::||||||||:   :  |:|| :| |:|  |:||||:||||||||||||:||  :| |||
Db   307 QARLDILKIHAGPITKHGEIDYEAIVKLSDGFNGADLRNVCTEAGMFAIRADHDFVVQED 366

Qy   395 FEMAVAKVM-KKETEKQMSLR                                        414
           | :||  |: |||  |  :  |
Db   367 FMKAVRKVADSKKLESKLDYK                                        387
```

```
RESULT 15
US-09-661-468-13
Sequence 13, Application US/09661468
; Patent No. 6376189
; GENERAL INFORMATION:
;  APPLICANT: FUJIWARA, TSUTOMU
;  APPLICANT: WATANABE, TAKESHI
;  APPLICANT: HORIE, MASATO
;  TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
;  TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
;  FILE REFERENCE: Q-53599
;  CURRENT APPLICATION NUMBER: US/09/661,468
;  CURRENT FILING DATE: 2000-09-13
;  PRIOR APPLICATION NUMBER: 09/055,699
;  PRIOR FILING DATE: 1998-04-07
;  PRIOR APPLICATION NUMBER: 08/820,170
;  PRIOR FILING DATE: 1997-03-19
;  PRIOR APPLICATION NUMBER: JP 63410/1996
;  PRIOR FILING DATE: 1996-03-19
;  PRIOR APPLICATION NUMBER: JP 69163/1997
;  PRIOR FILING DATE: 1997-03-05
;  NUMBER OF SEQ ID NOS: 95
;  SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 13
;  LENGTH: 389
;  TYPE: PRT
;  ORGANISM: Homo sapiens
US-09-661-468-13

Query Match             41.4%;  Score 877;  DB 4;  Length 389;
Best Local Similarity   48.0%;  Pred. No. 5e-74;
Matches  183;  Conservative   64;  Mismatches 128;  Indels    6;  Gaps    3;

Qy    35 YSLNIHEHQLLRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEVAVKVMGKN  94
            :  |||  ||         || :|  :|  :  |  ||| |||:||:|| |:  :
Db    12 YRKKLLEH---KEIDGRLKELREQLKELTKQYEKSENDLKALQSVGQIVGEVLKQLTEE   67

Qy    95 KVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKV 154
            || |   |||:|||   |  || |||||||||:    :|  |  |||||||:||    :
Db    68 KFIVKATNGPRVVGCRRQLDKSKLKPGTRVALDMTTLIMRYLPREVDPLVINMSHEDP  127

Qy   155 PDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAV 214
           :|::| :|||| :||:|:||||||| ::|||| :|||:||||||||||||||||||||
Db   128 GNVSYSEIGGLSEQIRELREVIELPLTNPELFQRVGIIPPKGCLLYGPPGTGKTLLARAV 187

Qy   215 AHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGS 274
           |  ||| |:|||::|||||||||:||:||:| : ||:|| |||||||||:||:||| ||:
Db   188 ASQLDCNFLKVVSSIVDKYIGESARLIREMFNYARDHQPCIIFMDEIDAIGGRRFSEGT  247

Qy   275 GNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPPTNE 334
             ::| :|||:|||||||:|||: : :|||:|||||: |||:|||||||:|||| | |||
Db   248 -SADREIQRTLMELLNQMDGFDTLHRVKVTMATNRPDTLDPALLRPGRLDRKIHIDLPNE 306

Qy   335 ESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHVTQED 394
           ::||||||||:   :  |:|| :| |:|  |:||||:||||||||||||:||  :| |||
Db   307 QARLDILKIHAGPITKHGEIDYEAIVKLSDGFNGADLRNVCTEAGMFAIRADHDFVVQED 366

Qy   395 FEMAVAKVM-KKETEKQMSLR                                        414
           | :||  |: |||  |  :  |
Db   367 FMKAVRKVADSKKLESKLDYK                                        387
```

Db   367   FMKAVRKVADSKKLESKLDYK 387

Search completed: November   9, 2003, 12:49:36
Job time : 29 secs

OM protein - protein search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run On: November 9, 2003, 12:48:17 ; Search time 78 Seconds
(without alignments)
920,400 Million cell updates/sec

Title: US-09-462-972-2
Perfect score: 2116
Sequence: 1 MALVGVELKHAAEGVPEANC...........VAKVMKKETEKNMSLRKLWK 418

Scoring table: BLOSUM62
GapOp 10.0 , Gapext 0.5

Searched: 644079 seqs, 171749292 residues
Total number of hits satisfying chosen parameters: 644079

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1:	/cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*	
2:	/cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*	
3:	/cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*	
4:	/cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*	
5:	/cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*	
6:	/cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*	
7:	/cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*	
8:	/cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*	
9:	/cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*	
10:	/cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*	
11:	/cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*	
12:	/cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*	
13:	/cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*	
14:	/cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*	
15:	/cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*	
16:	/cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*	
17:	/cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*	
18:	/cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*	644079

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	1670.5	78.9	402	9	US-09-925-301-984	Sequence 984, App
2	1540.5	72.8	401	12	US-10-032-585-7325	Sequence 7325, App
3	1508.5	71.3	389	15	US-10-128-714-3369	Sequence 3369, Ap
4	1507.5	71.2	389	15	US-10-128-714-8369	Sequence 8369, Ap
5	880	41.6	389	12	US-10-205-219-165	Sequence 165, Appl
6	877	41.4	389	10	US-09-976-165-13	Sequence 13, Appl
7	877	41.4	389	12	US-10-342-276-13	Sequence 13, Appl
8	874	41.3	451	15	US-10-128-714-8421	Sequence 8421, Ap
9	869	41.1	463	15	US-10-128-714-8256	Sequence 8256, Ap
10	868.5	41.0	439	15	US-10-128-714-3421	Sequence 3421, Ap
11	860	40.6	439	15	US-10-128-714-3522	Sequence 3522, Ap
12	860	40.6	428	15	US-10-128-714-8522	Sequence 8522, Ap
13	858.5	40.6	444	12	US-10-032-585-7385	Sequence 7385, Ap
14	855	40.4	465	12	US-10-032-585-7680	Sequence 7680, Ap
15	812	38.4	428	15	US-10-128-714-3418	Sequence 3418, Ap

16	809	38.2	464	15	US-10-128-714-8418	Sequence 8418, Ap
17	775	36.6	438	12	US-10-032-585-7464	Sequence 7464, Ap
18	769	36.3	465	15	US-10-128-714-3485	Sequence 3485, Ap
19	768	36.3	430	12	US-10-032-585-7453	Sequence 7453, Ap
20	758	35.8	421	15	US-10-128-714-8485	Sequence 8485, Ap
21	732.5	34.6	411	12	US-10-032-585-7284	Sequence 7284, Ap
22	708.5	33.5	359	15	US-10-128-714-3256	Sequence 3256, Ap
23	566.5	26.8	826	12	US-10-032-585-7227	Sequence 7227, Ap
24	539	25.5	664	15	US-10-156-761-12200	Sequence 12200, A
25	539	25.5	806	15	US-09-922-217-1118	Sequence 1118, Ap
26	533	25.2	806	14	US-10-025-380-1118	Sequence 1118, Ap
27	533	25.2	806	9	US-09-922-217-1117	Sequence 1117, Ap
28	532	25.1	806	14	US-10-025-380-1117	Sequence 1117, Ap
29	531.5	25.1	812	15	US-10-128-714-8092	Sequence 8092, Ap
30	531	25.0	791	15	US-10-128-714-3092	Sequence 3092, Ap
31	528.5	25.0	635	15	US-10-128-714-3092	Sequence 3092, Ap
32	524	24.8	413	11	US-09-809-391-751	Sequence 751, App
33	524	24.8	413	12	US-09-882-171-751	Sequence 751, App
34	523.5	24.7	616	14	US-10-047-260-30	Sequence 30, Appl
35	522	24.7	627	14	US-10-047-260-36	Sequence 36, Appl
36	521	24.6	698	11	US-09-828-062-9	Sequence 9, Appl
37	521	24.6	714	15	US-09-828-062-9	Sequence 9, Appl
38	520.5	24.6	760	10	US-09-712-363-292	Sequence 292, App
39	520.5	24.6	766	12	US-09-712-363-292	Sequence 292, App
40	518.5	24.5	628	9	US-10-032-585-7417	Sequence 7417, Ap
41	518.5	24.5	697	9	US-09-815-242-5794	Sequence 5794, Ap
42	516	24.4	644	9	US-09-815-242-12798	Sequence 12798, Ap
43	516	24.3	644	9	US-09-815-242-13760	Sequence 13760, A
44	515	24.3	644	9	US-09-815-242-12056	Sequence 12056, A
44	515	24.3	644	9	US-09-815-242-10318	Sequence 10318, A
45	515	24.3	644	11	US-09-948-137A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-925-301-984
Sequence 984, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 984
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-301-984

Query Match 78.9%; Score 1670.5; DB 9; Length 402;
Best Local Similarity 82.4%; Pred. No. 4.5e-145;
Matches 324; Conservative 37; Mismatches 31; Indels 1; Gaps 1;

```
Qy    26 KQGEGLRHYYSLNIHEHQLLRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSSYVG 85
             |||||||||||||||||||||||||||||:||||||:|||||||||:|||||||||||
Db    11 KAGSGLRQYYLSKIEELQLIVNDKSQNLRRLQAQRNELNARVRLLREELQLLQEQGSYVG 70

Qy    86 EVVKVNGKNKVLVKVHPEGKYVVVDIDXNIDITKITPSTRVALRNDSYVLHLVLPSKVDPL 145
             |||:|| ||||||||||||||||||:||||||||||||||||||||:|||||||||||
Db    71 EVVRAMDKKKVLVKVHPEGKFVVVDVDXNIDINDVTPNCRVALRNDSYTLRKILPNKVDPL 130

Qy   146 VNLMKVEKVPDSTYDMIGGLDQQIKEIKEVTELPIKHPELFESLGIAQPKGVLLYGPPGT 205
             |||||||:|||||||||||||||||||||||||||||||||||:||||||||||||||
Db   131 VSLMMVEKVPDSTYEMGGGLDKQIKEIKEVTELPVKHPELFEALGIAQPKGVLLYGPPGT 190
```

```
RESULT 2
US-10-032-585-7325
Sequence 7325, Application US/10032585
Publication No US20030180953A1
GENERAL INFORMATION:
APPLICANT: Terry, Roemer D.
APPLICANT: Bo, Jiang
APPLICANT: Charles, Boone
APPLICANT: Howard, Bussey
TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
FILE REFERENCE: 10182-005-999
CURRENT APPLICATION NUMBER: US/10/032,585
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 8000
SOFTWARE: Patentin version 3.1

SEQ ID NO 7325
LENGTH: 401
TYPE: PRT
ORGANISM: Candida albicans
US-10-032-585-7325

Query Match          72.8%;  Score 1540.5;  DB 12;  Length 401;
Best Local Similarity   76.3%;  
Matches   297;  Conservative   49;  Mismatches   42;  Indels   1;  Gaps   1;

Qy     30  GLRHYSLNIHEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEVVK   89
           |:|:|:||||||||:|:||||||:|:||||||||||||||||||||:|||||||||||||
Db     14  GIRPYFEQQIQDTEIRIQQCTTQNLRLEAQRNLNNKVRQLKDELKLLQEPGSYVGEVVK   73

Qy     90  VNGKNKVLVKVHPEGKYYVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLM  149
           ||:|||||||||||||||||||||||||:||:|:||||||||||||||:||||||||||:
Db     74  VMGLKKVLVKIHPEGKFIVNVTKDIDVKRLTPSIRVCLKADSHDLYKILPNKVDPLVSLM  133

Qy    150  KUEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTL  209
           ||:||||||||||||||||||||||:||||:|:||||||||||||:|||||||||||||||
Db    134  MVEKVPDSTYDMVGGLDKQIEKIEVIELPVKHPELFESLGIAQPKGVILYGPPGTKTL  193

Qy    210  LARAVAHHTDCTFIRVSGGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSAR  269
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db    194  LARAVAHHTECKFIRVSGGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSSR  253

Qy    270  MESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEF  329
           |:||||||||:||||||||||||||||:|||||||||:|||||||||||||||||||||:
Db    254  VE-GSSGGDSEVQRTMLELLNQLDGFESSKDIKIIMATNRDILDPALLRPGRIDRKIEF  312

Qy    330  TPTNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVH  389
           ||||||||||||||||||||||||||||||||:|||||||:|||||||||||||:||||
Db    313  PAPTVAARTDILKIHSRMNLTRGINLRKIAEKMNGCSGADVKGVCTEAGMYALRERRIH  372

Qy    390  VTQEDFEMAVAKVMKKETEKNMSLRKLWK  418
           ||||||||||:|||:||||||||||||||
Db    373  VTQEDFELAVAKVMSKNDDGAVSLQKLFK  401
```

RESULT 3
US-10-128-714-3369

```
Sequence 3369, Application US/10128714
Publication No. US20030119013A1
GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Hu, Wenqi
APPLICANT: Tishkoff, Daniel
APPLICANT: Zamudio, Carlos
APPLICANT: Eroshkin, Alexey M
APPLICANT: Lemieux, Sebastien M
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
TITLE OF INVENTION: Methods of Use
FILE REFERENCE: 10182-018-999
CURRENT APPLICATION NUMBER: US/10/128,714
CURRENT FILING DATE: 2002-04-23
PRIOR APPLICATION NUMBER: US 60/287,066
PRIOR FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: US 60/295,890
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/303,899
PRIOR FILING DATE: 2001-07-09
PRIOR APPLICATION NUMBER: US 60/316,362
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: US 60/285,697
PRIOR FILING DATE: 2001-04-23
NUMBER OF SEQ ID NOS: 8603
SOFTWARE: Patentin version 3.1

SEQ ID NO 3369
LENGTH: 389
TYPE: PRT
ORGANISM: Aspergillus fumigatus
US-10-128-714-3369

Query Match          71.3%;  Score 1508.5;  DB 15;  Length 389;
Best Local Similarity   77.8%;  
Matches   302;  Conservative   34;  Mismatches   51;  Indels   1;  Gaps   1;

Qy     31  LRHYSLNIHEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEVVK   90
           :|:||    :|    :|||||||||||||||||||||||||||||||||||||||||
Db      3  LDNYYRNKIESMKLEIIGCQAVLRRLEAQRNDYNSRVRLLREELGLLQQPGSYVGEVKV   62

Qy     91  MGRNKVLVKVHPEGKYYVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMK  150
           |:|:|||||||||||||||||:||||:|:|:||||||||||||||||||||||:|||:|:
Db     63  MSTKKVLVKVHPEGKYYVDIADGVDITKLTVGKRVALLSDSYKLEKMLPSSVDPLVSLMM  122

Qy    151  VEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLL  210
           |||||||||||||||||||||||:|||||:|||||||||||||||||||||||||||||
Db    123  VEKVPDSTYDMVGGLDQQIKEIKEVIELGLKHPELFESLGIAQPKGVLLYGPPGTGKTLL  182

Qy    211  ARAVAHHTDCTFIRVSGGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARM  270
           |||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183  ARAVAHHTDCRFIRVSGGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSSRI  242

Qy    271  ESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFP  330
           |:|||||||||||||||||||||||:||||||||:|:||||||:||:|||||||||:|||
Db    243  DS-AGSGDSEVQRTMLELLNQLDGFEPTKNIKIIMATNRLDIIDPALLRPGRIDRKIEFP  301

Qy    331  TPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHV  390
           ||||||||||||||||||||||||:|||||||||||||||||||||:||||||||||:||
Db    302  PPSVEARADILRIHSRSMNLTRGINLTKIAEKMNGCSGAELKGVCTEAGMYALRERRVHV  361

Qy    391  TQEDFEMAVAKVMKKETEKNMSLRKLWK  418
           ||||||:|||:|||:||||:|||:|:||
Db    362  TQEDFDLATAKILNKHDDKEVAVSKLFK  389
```

RESULT 4
US-10-128-714-8369

```
Sequence 8369, Application US/10128714
Publication No. US20030119013A1
GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Hu, Wenqi
APPLICANT: Tishkoff, Daniel
```

Left column (RESULT header area):

```
Qy    206  GKTLLARAVAHHTDCTFIRVSGGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSI  265
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    191  GKTLLARAVAHHTDCTFIRVSGGSELVQKYIGEGARMVRELFVMAREHAPSIIFMDEIDSI  250

Qy    266  GSARMESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDR  325
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    251  GSSRLEGGSG-GDSEVQRTMLELLNQLDGFEATKNIKVIMATNRIDILDSALLRPGRIDR  309

Qy    326  KIEFPTPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRE  385
           |||||:||||||||||||||||||||||||:|||||||||||||||:||||||||||||
Db    310  KIEFPPPNEEARLDILKIHSRKMNLTRGINLRKIAELMPGASGAEVKGVCTEAGMYALRE  369

Qy    386  RRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK  418
           |||||||||||:||:|||:::|||||||||||
Db    370  RRVHVTQEDFEMAVAKVMWQKDSEKNMSIKKLWK  402
```

APPLICANT: Zamudio, Carlos
APPLICANT: Eroshkin, Alexey M
APPLICANT: Lemieux, Sebastien M
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
TITLE OF INVENTION: Methods of Use
FILE REFERENCE: 10182-018-999
CURRENT APPLICATION NUMBER: US/10/128,714
CURRENT FILING DATE: 2002-04-23
PRIOR APPLICATION NUMBER: US/10/128,714
PRIOR FILING DATE: 2002-04-27
PRIOR APPLICATION NUMBER: US 60/287,066
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: US 60/303,899
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/295,890
PRIOR FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: US 60/295,697
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: US 60/316,362
PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 8603
SOFTWARE: Patentin version 3.1

SEQ ID NO 8369
LENGTH: 389
TYPE: PRT
ORGANISM: Aspergillus fumigatus

US-10-128-714-8369

Query Match 71.2%; Score 1507.5; DB 15; Length 389;
Best Local Similarity 77.8%; Pred. No. 4.2e-130;
Matches 302; Conservative 33; Mismatches 52; Indels 1; Gaps 1;

Qy   31 LRHYYSLNIHEHQLLRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEVKV   90
          ||||||||||||||||||||||||||||||||||:||||||| |||||| |||||||
Db   3 LDNYYRNKIESMKLEIIQGQAVLRRLEAQRNDYNSRVRLREELGLLQQPGSSYVGEVKV   62

Qy   91 MGKNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMK 150
          |||||||||||||||| |||||||||||||||||||||:|||||||||||||||||||
Db   63 MSTKKVLVKVHPEGKYVVDIADGVDITKLTVGKRVALLSDSYKLEKKMLPSSVDPLVSLMM 122

Qy   151 VEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESSLGIAQPKGVLLYGPPGTGKTLL 210
          ||:||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db   123 VEKVPDSTYDMIGGLDQQIKEIKEVIELGLKHPELFESSLGINQPKGVLLYGPPGTGKTLL 182

Qy   211 ARAVAHHTDCTFIRVSGGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSSARM 270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||::|
Db   183 ARAVAHHTDCRFIRVSGGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSSRI 242

Qy   271 ESGSGNDSEVQRTMLELLNQLLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFP 330
          :||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   243 DS-AGSGDSEVQRTMLELLNQLDGFEPTKNIKIIMATNRLDILDPALLRPGRIDRKIEFP 301

Qy   331 TPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHV 390
          ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db   302 PPRVEARADILRIHSRSRMNLTRGINLTKIAEKMNGCSGAELKGVCTEAGMYALRERRVHV 361

Qy   391 TOEDFEMAVAKVMKKETEKNMSLRKLMK 418
          |||||:|:||:|||||:|:||:||:|||
Db   362 TQEDFDLATAKILNKHDDKEVAVSKLFK 389

RESULT 5
US-10-205-219-165
Sequence 165, Application US/10205219
Publication No. US20030138803A1
GENERAL INFORMATION:
APPLICANT: Warner-Lambert Company
APPLICANT: Lee, Kevin
APPLICANT: Dixon, Alistair
APPLICANT: Brooksbank, Robert
APPLICANT: Pinnock, Robert
TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain
FILE REFERENCE: WL-A-018200
CURRENT APPLICATION NUMBER: US/10/205,219
CURRENT FILING DATE: 2002-07-24

PRIOR APPLICATION NUMBER: GB 0118354.0
PRIOR FILING DATE: 2001-07-27
NUMBER OF SEQ ID NOS: 197
SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 165
LENGTH: 389
TYPE: PRT
ORGANISM: Spermophilus tridecemlineatus
FEATURE:
OTHER INFORMATION: ATPase domain protein 44

US-10-205-219-165

Query Match 41.6%; Score 880; DB 12; Length 389;
Best Local Similarity 48.0%; Pred. No. 2.2e-72;
Matches 183; Conservative 65; Mismatches 127; Indels 6; Gaps 3;

Qy   35 YSLNIHEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSSYVGEVVKVMGKN   94
          |||||||||||| ||||||||||||||||||||||||||||||:|||||||:|: :| :
Db   12 YRKLLEH----KEIDGRLKELREQLKELTKQYEKSENDLKALQSVGQIVGEVLKQLTEE   67

Qy   95 KVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKV 154
          :|||||:|||||||||:||||||||||||||||:||:||:||||||:|||||| :||:
Db   68 KFIVKATNGPRYVVGCRRQLDKSKLKPGTRVALDMTILTIMRYLPREVDPLVYNMSHEDP 127

Qy   155 PDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESSLGIAQPKGVLLYGPPGTGKTLLARAV 214
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   128 GNVSYSEIGGLSEQIRELREVIELPLTNPELFQRVGIIPPKGCLLYGPPGTGKTLLARAV 187

Qy   215 AHHTDCTFIRVSGGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGS 274
          |||||| |||||||||::|||||||||||||||||||||||||||||||||||::||:|
Db   188 ASQLDCNFLKVVSSSIVDKYIGESARLIREMFNYARDHQPCIIFMDEIDAIGGRRFSEGT 247

Qy   275 GNGDSEVQRTMLELLNQLLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPTNE 334
          :|||||||||||||||||||||||:||||:|||||:|:||||||||:|||||||:::|:
Db   248 -SADREIQRTLMELLNQMDGFDTLHRVKNIMATNRPDTLDPALLRPGRLDRKIHIDLPNE 306

Qy   335 ESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHVTQED 394
          ::||||:|||||||||||||||||||||||||||||||||||||||:||||||||:|||
Db   307 QARLDIKIHAGPITKHGEIDYEAIVKLSDGFNGADLRNVCTEAGMFAIRADHDFVVQED 366

Qy   395 FEMAVAKVM-KKETEKNMSLR 414
          |:|||:|: ||:||:||:||
Db   367 FMKAVRKVADSKKLESKLDYK 387

RESULT 6
US-09-976-165-13
Sequence 13, Application US/09976165
Patent No. US20020107383A1
GENERAL INFORMATION:
APPLICANT: FUJIWARA, TSUTOMU
APPLICANT: WATANABE, TAKESHI
APPLICANT: HORIE, MASATO
TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
FILE REFERENCE: Q-53599
CURRENT APPLICATION NUMBER: US/09/976,165
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 09/565,538
PRIOR FILING DATE: 2000-05-05
PRIOR APPLICATION NUMBER: 09/055,699
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: 08/820,170
PRIOR FILING DATE: 1997-03-19
PRIOR APPLICATION NUMBER: JP 63410/1996
PRIOR FILING DATE: 1996-03-19
PRIOR APPLICATION NUMBER: JP 69163/1997
PRIOR FILING DATE: 1997-03-05
NUMBER OF SEQ ID NOS: 95
SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 13
LENGTH: 389
TYPE: PRT

```
; ORGANISM: Homo sapiens
US-09-976-165-13

  Query Match            41.4%;  Score 877;  DB 10;  Length 389;
  Best Local Similarity  48.0%;  Pred. No. 4.1e-72;
  Matches 183; Conservative   64; Mismatches 128; Indels   6; Gaps   3;

Qy   35 YSLNIHEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEVVKVMGKN 94
         ::|::|| |||::|:|:::::|| || ::::|| |:::||||| ||:|:||||:|:|:|
Db   12 YRKLLEH----KEIDGRLKELRQLKELTKQYEKSENDLKALQSVGQIVGEVLKQLTEE 67

Qy   95 KVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKV 154
         ||||:||||||||||||||||||||||||:||||||:|| |||||||||||:||||||:
Db   68 KFIVKATNGPRYVVGCRRQLDKSKLKPGTRVALDMTLTIMRYLPREVDPLVYNMSHEDP 127

Qy  155 PDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAV 214
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  128 GNVSYSEIGLSEQIRELREVIELPLTNPELFQRVGIIPPKGCLLYGPPGTGKTLLARAV 187

Qy  215 AHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGS 274
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  188 ASQLDCNFLKVVSSSIVDKYIGESARLIREMFNYARDHQPCIIFMDEIDAIGGRRFSEGT 247

Qy  275 GNGDSEVQRTMLELLNQLLGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPPTNE 334
         :||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db  248 -SADREIQRTLMELLNQMDGFDTLHRVKMTMATNRPDTLDPALLRPGRLDRKIHIDLPNE 306

Qy  335 ESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHVTQED 394
         ::|||||||||||||||||:||:|||||||:||||||||||||||:|||||||||||||
Db  307 QARLDILKIHAGPIITKHGEIDYEAIVKLSDGFNGADLRNVCTEAGMFAIRADHDFVVQED 366

Qy  395 FEMAVAKVM--KKETEKNMSLR 414
         ||||||||   ::  ::|::|:
Db  367 FMKAVRKVADSKKLESKLDYK 387
```

```
RESULT 7
US-10-342-276-13
; Sequence 13, Application US/10342276
; Publication No. US20030143688A1
; GENERAL INFORMATION:
; APPLICANT: FUJIWARA, TSUTOMU
; APPLICANT: WATANABE, TAKESHI
; APPLICANT: HORIE, MASATO
; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
; TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
; FILE REFERENCE: Q-53599
; CURRENT APPLICATION NUMBER: US/10/342,276
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: US/09/976,165
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 09/565,638
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: 09/055,699
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: 08/820,170
; PRIOR FILING DATE: 1997-03-19
; PRIOR APPLICATION NUMBER: JP 63410/1996
; PRIOR FILING DATE: 1996-03-19
; PRIOR APPLICATION NUMBER: JP 69163/1997
; PRIOR FILING DATE: 1997-03-05
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 13
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Homo sapiens

US-10-342-276-13

  Query Match            41.4%;  Score 877;  DB 12;  Length 389;
  Best Local Similarity  48.0%;  Pred. No. 4.1e-72;
  Matches 183; Conservative   64; Mismatches 128; Indels   6; Gaps   3;
```

```
Qy   35 YSLNIHEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEVVKVMGKN 94
         ::|::|| |||::|:|:::::|| || ::::|| |:::||||| ||:|:||||:|:|:|
Db   12 YRKLLEH----KEIDGRLKELRQLKELTKQYEKSENDLKALQSVGQIVGEVLKQLTEE 67

Qy   95 KVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKV 154
         ||||:||||||||||||||||||||||||:||||||:|| |||||||||||:||||||:
Db   68 KFIVKATNGPRYVVGCRRQLDKSKLKPGTRVALDMTLTIMRYLPREVDPLVYNMSHEDP 127

Qy  155 PDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAV 214
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  128 GNVSYSEIGLSEQIRELREVIELPLTNPELFQRVGIIPPKGCLLYGPPGTGKTLLARAV 187

Qy  215 AHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGS 274
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  188 ASQLDCNFLKVVSSSIVDKYIGESARLIREMFNYARDHQPCIIFMDEIDAIGGRRFSEGT 247

Qy  275 GNGDSEVQRTMLELLNQLLGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPPTNE 334
         :||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db  248 -SADREIQRTLMELLNQMDGFDTLHRVKMTMATNRPDTLDPALLRPGRLDRKIHIDLPNE 306

Qy  335 ESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHVTQED 394
         ::|||||||||||||||||:||:|||||||:||||||||||||||:|||||||||||||
Db  307 QARLDILKIHAGPIITKHGEIDYEAIVKLSDGFNGADLRNVCTEAGMFAIRADHDFVVQED 366

Qy  395 FEMAVAKVM--KKETEKNMSLR 414
         ||||||||   ::  ::|::|:
Db  367 FMKAVRKVADSKKLESKLDYK 387
```

```
RESULT 8
US-10-128-714-8421
; Sequence 8421, Application US/10128714
; Publication NO. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Broshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: Patentin version 3.1
;
; SEQ ID NO 8421
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus

US-10-128-714-8421

  Query Match            41.3%;  Score 874;  DB 15;  Length 451;
  Best Local Similarity  45.4%;  Pred. No. 9.7e-72;
  Matches 189; Conservative   72; Mismatches 139; Indels   16; Gaps   5;
```

```
Qy    9 KHAAEGVPEANCSAKPTKQG-------EGLRHYYSLNIHEHQLLLRQKTHNLNRLEAQRN 61
         ||  ||  |:||| ||||:|        |::|:||||||||||||||||||||||||||
Db   44 QHFFEALPEG--FAPPTWNGGSDPEREQALEDY----KKSLLELREWEAKLLKNLRMGIK 96

Qy   62 DLNSRVRMLREELQLLQEPGSYVGEVVKVMGKNKVLVKVHPEGKYVVDIDKNIDITKITP 121
         ||  |:|:|:||:||||||||||||||||||     ::|||||||||||||||||||||
```

```
Db   97 DLQREFDISEENIKALQSVGQIIGEVLKQLDEERVIVKASSGPRYVVGCRSKVDKSKLKQ 156
Qy  122 STRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIK 181
Db  157 GTRVALDMTLTIMRMLPREVDPLVVNMSLEDPGQISPAGIGGLNDQIRELREVELPLK 216
Qy  182 HPELFFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKYIGEGSRM 241
Db  217 NPELFQRVGIKPPKGVLLYGPPGTGKTLLARAVASSMETNFLKVVSSAIVDKYICESARL 276
Qy  242 VRELFVWAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNKI 301
Db  277 IREMFGYAKEHPCIIFMDEIDAIGGRRFSEGT-SADREIQRTLMELLNQLDGFDYLGKT 335
Qy  302 KVLMATNRIDILDQALLRPGRIDRKIEFPTPNEESRLDILKIHSRRMNLMRGIDLKKIAE 361
Db  336 KIIMATNRPDTLDPALLRAGRLDRKIEIPLPNEVGRLEILKIHSSTVQLEGEIDFESVVK 395
Qy  362 KMMGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMK-KETEKRMSLRKL 416
Db  396 MSDGLNGADLRNVVTERGLFAIKDYRDAISQODFNKAVRKVAEAKLLEGKLEYQKL 451
```

RESULT 9
US-10-128-714-8256
Sequence 8256, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 8256
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8256

Query Match   41.1%; Score 869; DB 15; Length 463;
Best Local Similarity   43.7%; Pred. No. 2.9e-71;
Matches   177; Conservative   82; Mismatches   124; Indels   22; Gaps   4;

```
Qy   16 PEANCSAKPTKQGEGLRHYSLNIHEHQLLRQKTHNLNRL----------EAQRNDLNS 65
Db   60 PTSRCK-------LRYLRMQRVHDHLLEEEVENMERLRKAKAQATLDSVSRGDLDI 110
Qy   66 RVRMLREELQLLQEPGS--YGEVVKVWGKNKVLVKVHPEGKYVVDIDKNIDITKITPST 123
Db  111 MDRNADERSVDDMRGSPMGVGNLELLIDDDHAIVSSATGPEYYYSIMSFVDKDLLEPGA 170
Qy  124 RVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHP 183
Db  171 SILLHHKSVSVUGVLTEESDPLVSVNMKLDKAPTESYADIGGLESQIEVRESVLPLLHP 230
```

RESULT 10
US-10-128-714-3421
Sequence 3421, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3421
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-3421

Query Match   41.0%; Score 868.5; DB 15; Length 393;
Best Local Similarity   47.5%; Pred. No. 2.5e-71;
Matches   182; Conservative   71; Mismatches   127; Indels   3; Gaps   3;

```
Qy  184 ELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKYIGEGSRMVR 243
Db  231 ELYEEMGIKPPKGVLLYGAPGTGKTLLAKAVANQTSAFFLRIVSGSELIQKYLGDGPRLVR 290
Qy  244 ELFVWAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNKIKV 303
Db  291 QIFQVAAEHAPSIVFIDEIDAIGTKRYDSTSG--GEREIQRTMLELLNQLDGFDDRGQVKV 349
Qy  304 LMATNRIDILDQALLRPGRIDRKIEFPTPNEESRLDILKIHSRRMNLMRGIDLKKIAEKM 363
Db  350 IMATNKIETLDPALIRPGRIDRKILFENPDENTKKKIFTLHTSKMSLADDVDLDEFINQK 409
Qy  364 NGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKETE 408
Db  410 DDLSGADIRAICTEAGLMALRERRMRVQMDDFRAAREKIMKTKQD 454

Qy   36 SLNIHEHQLL-LRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEVVKVWGKN 94
Db   12 ALEDYKKSLLELRWEAKLKNLRMGIKDLQREFDISEENIKALQSVGQIIGEVLKQLDEE 71
Qy   95 KVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKV 154
Db   72 RFIVKASSGPRYVVGCRSKVDKSKLKQGTRVALDMTLTIMRMLPREVDPLVVNMSLEDP 131
Qy  155 PDSTYDMIGGLDQQIKEIKEVIELPIKHPELFFESLGIAQPKGVLLYGPPGTGKTLLARAV 214
Db  132 GQISPAGIGGLNDQIRELREVELPLKNPELFQRVGIKPPKGVLLYGPPGTGKTLLARAV 191
Qy  215 AHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVWAREHAPSIIFMDEIDSIGSARMESGS 274
Db  192 ASSMETNFLKVVSSAIVDKYIGESARLIREMFGYAKEHEPCIIFMDEIDAIGGRRFSEGT 251
Qy  275 GNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPTPNE 334
```

RESULT 11
US-10-128-714-3522
; Sequence 3522, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3522
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-3522

Query Match   40.6%; Score 860; DB 15; Length 439;
Best Local Similarity   47.8%; Pred. No. 1.8e-70;
Matches 183; Conservative 61; Mismatches 113; Indels 26; Gaps 3;

Qy   53 LNRLEAQRNDLNSRVR------------------MLREELQLLQEPGSYVGEVVKVMG- 92
               ||::| ::| |::| ||| : |:|| ||
Db   49 LKKLEKQIKDKQASVNEKIGVKESDTGLAPPHLWDVAADRQRMAEEQPLQVARCTKIISD 108

Qy   93 ----KNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVN 147
                 ||::| ||:|| || : || |||||:||||:||:| |:|:| ||| |:|||
Db   109 EKDPDKSKYVINVKQIAKRVVNLGERVSPTDIEEGMRVGVDRNKYQIMLPLPPKIDPSVT 168

Qy   148 LMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESSLGIAQPKGVLLYGPPGTGK 207
               |:|| :|||||||||| ||:| |||:| || | ||||| | :: || || ||
Db   169 MMTVEDKPDVTYGDVGGCKEQIEKLREVVEMPLLSPERFVNLGIDPPKGALLYGPPGTGK 228

Qy   208 TLLARAVAHHTDCTFIRVSGSELVQKYVGEGARMVRELFVMAREHAPSIIFMDEIDSIGS 267
               ||||| | | ||||||| ||:|||||| | |||||| | ||||:||||||||||
Db   229 TLCARAVANRIDATFIRVISSELVQKYVGEGARMVRELFEMARTKKACIIFFDEIDAVGG 288

Qy   268 ARMESGSGNGDSEVQORTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKI 327
               |||||||||||:|||||||||||||||||| |||:| :| |||: | |||:||||||
Db   289 ARFDGAG--GDNEVQRTMLELITQLDGPARGNIKVMFATNRPSTLDPALMRPGRIDRKI 347

Qy   328 EFPTPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERR 387
               ||| | ||||| |||| |||:| |||| ||:||||||||||| ||||:||||:|
Db   348 EFSLPDVEGRANILRIHAKSMSVERDIRWELISRLCPNATGAELRSVATEAGMFAIRARR 407

Qy   388 FNKAVRKVAEAKLEGKLEYQKL 393
               |:|| :| ||::| || | |
Db   408 KVATEKDFLAAVDKVIKGNLKFN 430

RESULT 12
US-10-128-714-8522
; Sequence 8522, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 8522
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8522

Query Match   40.6%; Score 860; DB 15; Length 439;
Best Local Similarity   47.8%; Pred. No. 1.8e-70;
Matches 183; Conservative 61; Mismatches 113; Indels 26; Gaps 3;

Qy   53 LNRLEAQRNDLNSRVR------------------MLREELQLLQEPGSYVGEVVKVMG- 92
               ||::| ::| |::| ||| : |:|| ||
Db   49 LKKLEKQIKDKQASVNEKIGVKESDTGLAPPHLWDVAADRQRMAEEQPLQVARCTKIISD 108

Qy   93 ----KNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVN 147
                 ||::| ||:|| || : || |||||:||||:||:| |:|:| ||| |:|||
Db   109 EKDPDKSKYVINVKQIAKRVVNLGERVSPTDIEEGMRVGVDRNKYQIMLPLPPKIDPSVT 168

Qy   148 LMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESSLGIAQPKGVLLYGPPGTGK 207
               |:|| :|||||||||| ||:| |||:| || | ||||| | :: || || ||
Db   169 MMTVEDKPDVTYGDVGGCKEQIEKLREVVEMPLLSPERFVNLGIDPPKGALLYGPPGTGK 228

Qy   208 TLLARAVAHHTDCTFIRVSGSELVQKYVGEGARMVRELFVMAREHAPSIIFMDEIDSIGS 267
               ||||| | | ||||||| ||:|||||| | |||||| | ||||:||||||||||
Db   229 TLCARAVANRIDATFIRVISSELVQKYVGEGARMVRELFEMARTKKACIIFFDEIDAVGG 288

Qy   268 ARMESGSGNGDSEVQORTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKI 327
               |||||||||||:|||||||||||||||||| |||:| :| |||: | |||:||||||
Db   289 ARFDGAG--GDNEVQRTMLELITQLDGPARGNIKVMFATNRPSTLDPALMRPGRIDRKI 347

Qy   328 EFPTPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERR 387
               ||| | ||||| |||| |||:| |||| ||:||||||||||| ||||:||||:|
Db   348 EFSLPDVEGRANILRIHAKSMSVERDIRWELISRLCPNATGAELRSVATEAGMFAIRARR 407

Qy   388 VHVTQEDFEMAVAKYMKKETEKN 410
               |:||:||| | ||:| ||| | |
Db   408 KVATEKDFLAAVDKVIKGNLKFN 430

RESULT 13

US-10-032-585-7385
; Sequence 7385, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 7385
; LENGTH: 444
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7385

Query Match   40.6%; Score 858.5; DB 12; Length 444;
Best Local Similarity   45.0%; Pred. No. 2.5e-70;
Matches 189; Conservative 69; Mismatches 125; Indels 37; Gaps 6;

```
Qy   17 FANCSAKPTKQGEG------LRHYSLNIHEHQLLLRQKTHNILNRLEAQRNDLNSRVR----  68
Db   15 EEETSEKITPLTEGDIQVLKHYGSAPYADS---UKEIDKDLKPIE-ERIKTNQGIQESDT   70

Qy   69 -------MLRELQLLLQEPGSYVGEVVKVM------------------------GKNVLVKVHPEG 104
Db   71 GLAPPHLWDVLGDKQRMOEEQSLQVARCTKIIEASQPQPTGLQNADTKSKYVINIKOIA  130

Qy  105 KYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGG 164
Db  131 KFVVGLGERLSPTDIEEGMRVGVDRHKYEIQLPLPPRIDPSVTMMIVEEKPDVTYSDVGG 190

Qy  165 LDQQIKEIKEVTELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIR 224
Db  191 CKEQIEKLREVVELPLLSPERFVKLGIDPPKGILLGPPGTGKTLCARAVANRTDATFIR  250

Qy  225 VSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRT 284
Db  251 VIGSELVQKYVGEGARMVRELPEMARTKKACIIFPDEVDAIGGARFPDGAG-GDNEVQRT 309

Qy  285 MLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPTPNEESRLDIKIH  344
Db  310 MLELITQLDGPDPRGNIKVMFATNRNTLDPALLRPGRIDRKVEFSLPDLEGRANIFRIH  369

Qy  345 SRRMNLMRGIDLKKIAEKMNGASGAEL.KAVCTEAGMFALRERRVHVTQEDFEMAVAKVMK 404
Db  370 SKTMSVEKDIRWELISRLCPNATGAELRSVCTEAGMFAIRARRKVANEKDFLKAVDKVIK  429
```

RESULT 14
US-10-032-585-7680
; Sequence 7680, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 7680
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7680

Query Match   40.4%; Score 855; DB 12; Length 465;
Best Local Similarity   44.7%; Pred. No. 5.7e-70;
Matches 177; Conservative 77; Mismatches 128; Indels 14; Gaps 3;

```
Qy   16 PEANCSAKPTKQGEGLRHYYSLNIHEHQLLLRQKTHNILNRLEAQRNDLNSRVRMLREELQ  75
Db   55 PNTRCKLKLKLE-------RIKDHLLLEEBFVTNQEAFQ---PTEARQAEEREKVD      101

Qy   76 LLQEPGSYVGEVVKVMGKNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLH  135
Db  102 EURGYPMAIGTLEEIIDDDHAIVSSTASSEYYVSIMSFVDKGLLEPGCSVLLHHKTVAVV  161

Qy  136 LVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPK  195
Db  162 GVLQDDADPMVSVMKLDKSPTSSYADIGGLESQIOEIKESVELPLTHPELYEEMGIKPPK  221

Qy  196 GVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGGSELVQKYIGEGSRMVRELFVMAREHAPS 255
Db  222 GVILYGAPGTGKTLLAKAVANQTSATFLRIVSSELIQKYLGDGPRLCRQIFQIAADHAPS  281

Qy  256 IIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQ  315
Db  282 IVFIDEIDAIGITKRYESTSG-GEREIQRTMLELLNQDGFDDRGDIKVIMATNKIESLDP  340

Qy  316 ALLRPGRIDRKIEFPTPNEESRLDIKIHSRSMNLMRGIDLKKIAEKMNGASGAELKAVC   375
Db  341 ALIRPGRIDRKILFENEDANTKKKILIIHTSKMSLADDVNLDEIVTGKDDLSGADIKAIC  400

Qy  376 TEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNM  411
Db  401 TEAGLLALRERRMQVKAEDFKSAKERVLKNKVEENL  436
```

RESULT 15
US-10-128-714-3418
; Sequence 3418, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3418
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-3418

Query Match   38.4%; Score 812; DB 15; Length 428;
Best Local Similarity   43.8%; Pred. No. 4.5e-66;
Matches 171; Conservative 78; Mismatches 117; Indels 24; Gaps 3;

```
Qy   46 LRQKTHNILNRLEAQRNDLNSRVRMLREELQLLLQEPGSYVGEVVKVM-------------  91
```

```
Db    39  LRIMKSEYQRLTHEQNTMREKVRDNQEKIENRQLPYLVGNVVELLDLDVEAEAEEGAN  98
              ||       :||:    ::|         :||   ||      ||  :||:::
Qy    92  ---------GKNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVD  143
              ||:|:|                         |:                ::
Db    99  IDLDATRVGKSAV-IKTSTRQTIYLPLIGLVDHEKLKPGDLIGVNKDSYLILDTLPAEYD  157
              |||||                ||:|:|||:     ||||
Qy   144  PLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIIKHPELFESLGIIAQPKGVLLYGPP  203
              :||:|           ::||:||||||   |:|    |||:|||:||::||||
Db   158  NRVKAMEVDEKPTEKYTDIGGLDKQIEEIVEAIVWPMKEAERFTKIGIKAPKGALMYGPP  217
              :|||||||                  |    ||:::    :|||:|:||
Qy   204  GTGKTLLARAVAHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEID  263
              |||||||||        |||:::|  ||:|  |  ::|  ||    |||||||:
Db   218  GTGKTLLARACAAETNATFLKLAGPQLVQMFIGDGAKLVRDCFALAKEKAPSIIFIDELD  277
              ::                 |:  |  ::|         |:|:|:|||   ::::|
Qy   264  SIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRI  323
              :|:|      ||  ||||:|||||||:||||::     |:|:|||||:|:||
Db   278  AVGTKRFDS-EKSGDREVQRTMLELLNQLDGFASDDRIKVLAATNRIDVLDPALLRSGRL  336
              ||:|||||   |:||::::|       |::|||:|    :::|:|||:||::|:
Qy   324  DRKIEFPTPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFAL  383
              ||||:|| |||:||:|||:||||||||||:||:||: |:|    :||:||:||||:|
Db   337  DRKIEFPLPNEEARANILQIHSRKMAVDDAVNWAELARSTDEFGGAQLKAVCVEAGMIAL  396
              |  |  ||                :|     :|  |:|| |||: |::|
Qy   384  RERRVHVTQEDFEMAVAKVMKKETEKNMSL  413
              |:     |   :|: |   |:::
Db   397  RKGMSKVGHENYVDAIAEVQAKKKDTNMGI  426
```

OM protein - protein search, using sw model

GenCore version 5.1.6

Run on: November 9, 2003, 12:36:11 ; Search time 44 Seconds
(without alignments)
913.602 Million cell updates/sec

Title: US-09-462-972-2
Perfect score: 2116
Sequence: 1 MALVGVELKHAAEGVPEANC........VAKVMKKETEKNMSLRKLWK 418

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:*
1: piri:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	Match ID	Description
1	1680	79.4	406	2	S61923	SUG1 protein - mou
2	1675	79.2	406	2	S65536	26S proteasome cha
3	1665	78.7	406	2	S60343	Tripl protein - hu
4	1630	77.0	389	2	JN0610	probable transcrip
5	1624.5	76.8	374	2	T43799	probable transcrip
6	1597	75.5	403	2	S45176	26S protease reg
7	1584.5	74.9	435	2	S51042	tat-binding reg
8	1572	74.3	416	2	T27048	hypothetical protei
9	1528.5	72.2	389	2	T49402	probable 26S prote
10	1504.5	71.1	405	2	S64052	26S proteasome reg
11	1482.5	70.1	443	2	T33633	hypothetical prote
12	1076	50.9	400	2	F90131	26S proteasome SU
13	955.5	45.2	399	2	G71242	probable 26S prote
14	937.5	44.3	399	2	H75207	ATP-dependent 26S
15	929	43.9	398	2	G69496	ATP-dependent 26S
16	923	43.6	283	2	S52928	XSUG1 protein - Af
17	923	43.6	410	2	C69197	ATP-dependent 26S
18	906	42.8	430	2	G64446	ATP-dependent 26S
19	903	42.7	443	2	T31800	hypothetical prote
20	901	42.6	440	2	A44468	26S proteasome reg
21	897	42.4	440	2	S74197	ATP-dependent 26S
22	894.5	42.3	426	2	G96577	26S proteasome ATP
23	892	42.2	426	2	T09104	26S proteasome ATP
24	888	42.0	433	1	S24353	proteasome 26S sub
25	888	42.0	433	2	S53709	MSS1 protein homol
26	880	41.6	435	2	T20152	26S proteasome CAD
27	878.5	41.5	435	2	JC5349	hypothetical prote
28	877	41.4	389	2	S71316	proteasome endopep
29	865.5	40.9	451	2	H96577	hypothetical prote
30	864	40.8	398	2	T32268	hypothetical prote
31	859.5	40.6	443	2	E84585	26S proteasome sub
32	859.5	40.6	448	2	S39348	26S ATP/ubiquitin-
33	858.5	40.6	443	2	T08959	proteinase homolog
34	856.5	40.5	471	2	T48743	probable 26S ATP/t
35	849.5	40.1	433	2	S39349	MSS1 protein homol
36	854	40.4	448	2	T03776	tat binding protei
37	854	40.4	439	2	T49507	probable 26S prote
38	852	40.3	440	2	T39558	26S proteinase reg
39	849.5	40.1	467	2	S46631	26S proteasome reg
40	848	40.1	437	2	F72504	probable 26S prote
41	817	38.6	409	2	S67156	26S proteasome reg
42	816.5	38.6	437	2	S67156	26S proteasome reg
43	811	38.3	493	2	D87802	protein C10G11.8 (
44	808.5	38.2	404	2	A34832	Tat-binding protei
45	807	38.1	394	2	E90115	26S protease regul

ALIGNMENTS

RESULT 1
S61923
SUG1 protein - mouse
SUG1 protein - mouse (Mus musculus (house mouse))
C;Species: Mus musculus (house mouse)
C;Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 19-Jan-2001
C;Accession: S61923
R;von Baur, E.; Zechel, C.; Heery, D.; Heine, M.J.S.; Garnier, J.M.; Vivat, V.; Le Dou
EMBO J. 15, 110-124, 1996
A;Title: Differential ligand-dependent interactions between the AF-2 activating domain
A;Reference number: S61923; MID:9617766; PMID:8598193
A;Accession: S61923
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-406 <VOM>
A;Cross-references: EMBL:Z54219; NID:g1200105; PIDN:CAA90961.1; PID:g1165125
C;Superfamily: ATP-dependent 26S proteinase; FtsH/SEC18/CDC48-type ATP-binding domain
C;Keywords: ATP; nucleotide binding; P-loop
F;163-373/Domain: FtsH/SEC18/CDC48-type ATP-binding domain homology <VATP>
F;190-197/Region: nucleotide-binding motif A (P-loop)

Query Match 79.4%; Score 1680; DB 2; Length 406;
Best Local Similarity 79.2%; Pred. No. 7.2e-99;
Matches 331; Conservative 37; Mismatches 38; Indels 12; Gaps 2;

```
Qy    1 MALVGVELKHAAEGVPEANCSAKPTKQGEGIJRHYYSLNIHEHQLLLRQKTHLNRLEAQR 60
          ||| || : :| | :                :||| | || || :||| || ||| |||
Db    1 MALDGPEQMELEEG----------KAGSGLRQYYLSKIEELQLIVNDKSQNLRRLQAQR 49

Qy   61 NDLNSRVRMLREELQLLQEPGSSYVGEVKVKMGKNKVLVKVHPEGKYXVVDIKDNIDITKIT 120
          | ||: ||:||||||||| |||||| | : ::||||||||||||| :||: :|||| :|
Db   50 NELNAKVRLLREELQLLQEQGSSYVGVVRAMDKKKVLVKVHPEGKFVVDVDXNIDINDVT 109

Qy  121 PSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDDQQIKEIKEVIELPI 180
          |: :||||||||:|| :||:||||| :|| ||||||||:|||||| |||| ||||||||:
Db  110 PNCRVALRNDSYTLHKILPNKVDPLVSLMMVEKVPDSTYEMIGGLDKQIKEIKEVIELPV 169

Qy  181 KHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKYIGEGSR 240
          |||||| |||||||||||||||||||||||||||||||||||||||||||||| ||||:|
Db  170 KHPELFEALGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKFIGEGAR 229

Qy  241 MVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNK 300
          ||||||||||||||||||||||||||| ::  ||| |||||||||||||||||||||::
Db  230 MVRELFVMAREHAPSIIFMDEIDSIGSSRLEGGSG-GDSEVQRTMLELLNQLDGFEATKN 288

Qy  301 IKVLMATNRIDILDLDQLLRPGRIDRKIEFPPNEESRLDILKIHSRKMNLMRGIDLKKIA 360
          |||:|||||||||| |:||||||||||||||| | ||||||||||||||| ||| |:||
Db  289 IKVIMATNRIDILDSALLRPGRIDRKIEFPPPNEBARLDILKIHSRKMNLTRGINLRKIA 348

Qy  361 EKMNGHASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
          | |  :||||::| ||||||| ||||||||||||||||||||||::|:|||||:||||
Db  349 ELMPGASGAEVKGVCTEAGMYALRERRVHVTQEDFEMAVAKVMQKDSEKNMSIRKLWK 406
```

RESULT 2

```
S65536
26S proteasome chain p45 - human
C;Species: Homo sapiens (man)
C;Date: 19-Mar-1997 #sequence_revision 09-May-1997 #text_change 19-Jan-2001
A;Accession: S65536; S65488; S65536; 153510
R;Akiyama, K.; Yokota, K.; Kagawa, S.; Shimbara, N.; DeMartino, G.N.; Slaughter, C.A.; N
 FEBS Lett. 363, 151-156, 1995
A;Reference number: 153510; MUID:95246863; PMID:7729537
A;Title: cDNA cloning of a new putative ATPase subunit p45 of the human 26S proteasome,
A;Accession: S65536
A;Molecule type: mRNA
A;Residues: 1-406 <AKI>
A;Cross-references: EMBL:D44467; NID:g976226; PIDN:BAA07919.1; PID:g976227
A;Accession: S65488
A;Molecule type: protein
A;Residues: 28-36,'X',38;58-60,'X',62-77;95-111;126-135,'X',137-145,'X',148;163-182;240
A;Keywords: ATP; nucleotide binding; P-loop
A;Superfamily: ATP-dependent 26S proteinase; FtsH/SEC18/CDC48-type ATP-binding domain ho
A;Genetics:
A;Gene: GDB:PSMC5; P45; S8; TRIP1
A;Cross-references: GDB:581712; OMIM:601681
A;Map position: 17q24-17q25
A;Superfamily: ATP-dependent 26S proteinase; FtsH/SEC18/CDC48-type ATP-binding domain ho
F;163-373/Domain: FtsH/SEC18/CDC48-type ATP-binding domain homology <VATP>
F;190-197/Region: nucleotide-binding motif A (P-loop)

Query Match       79.2%;  Score 1675;  DB 2;  Length 406;
Best Local Similarity   78.9%;
Matches 330;  Conservative 37;  Mismatches 39;  Indels 12;  Gaps 2;

Qy     1 MALVGVELKHAAEGVPEANCSAKPTKQGSGLRHYSLNIHEHQLLLRQKTHNLNRLEAQR 60
         |  :|::| ||  |:                  || :||::|||| ||||||||||||||
Db     1 -------------------------------KAGSGLRQYLSKIEELQLIVNDKSQNLRLQAQR 49

Qy    61 NDLNSRVRMLREELQLLQEPGSYVGEVVKVKMGKNKVLVKVHPEGKYVVDIDKNIDITKIT 120
         |:||::||:||.ELQLLQE.GSYVGEVV:   .:.KVLVKVHPEGK:VVD:DKNID.D.T
Db    50 NELNAKVRLLRRELQLLQEQGSYVGEVVRAMDKKKVLVKVHPEGKFVVDVDKNIDNDVT 109

Qy   121 PSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPI 180
         |. RVALRNDSY.LH :LP.KVDPLV.LM VEKVPDSTY:MIGGLD:QIKEIKEVIELP:
Db   110 PNCRVALRNDSYTLHKILPNKVDPLVSLMMVEKVPDSTYEMIGGLDKQIKEIKEVIELPV 169

Qy   181 KHPELFEALGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKFIGEGAR 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   170 KHPELFEALGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKFIGEGAR 229

Qy   241 MVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNK 300
         |||||||||||||||||||||||||||| ::   |.GDSEVQRTMLELLNQLDGFEA::|
Db   230 MVRELFVMAREHAPSIIFMDEIDSIGSSRLEGGSG--GDSEVQRTMLELLNQLDGFEATKN 288

Qy   301 IKVLMATNRIDILDQALLRPGRIDRKIEFPPTPNEESRLDILKIHSRRMNLMRGIDLKKIA 360
         ||| |||||||||| |||||||||||||||| :||:.|||||||||| |||:||| |:||
Db   289 IKVIMATNRIDILDSALLRPGRIDRKIEFPPNEEARLDILKIHSRKMNLTRGINLRKIA 348

Qy   361 EKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
         | | |||||::|.||||||||:||||||||||||| ||||||||:|::|||||:|||||
Db   349 ELMPGASGAEVKGVCTEAGMYALRERRVHVTQEDEEMAVAKVMQKDSEKNMSIKKLWK 406
```

RESULT 3

```
S60343
TrIp1 protein - human
C;Species: Homo sapiens (man)
C;Date: 06-Sep-1996 #sequence_revision 27-Feb-1997 #text_change 19-Jan-2001
A;Accession: S60343
R;Lee, J.W.; Ryan, F.; Swaffield, J.C.; Johnston, S.A.; Moore, D.D.
 Nature 374, 91-93, 1995
A;Title: Interaction of thyroid-hormone receptor with a conserved transcriptional mediat
A;Reference number: S60343; MUID:95174891; PMID:7870181
A;Accession: S60343
A;Status: Preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A;Molecule type: mRNA
```

RESULT 4 (right column)

```
A;Residues: 1-406 <LEE>
A;Cross-references: GB:L38810; NID:g6953369; PIDN:AAC41735.1; PID:g6953370
C;Superfamily: ATP-dependent 26S proteinase; FtsH/SEC18/CDC48-type ATP-binding domain
C;Keywords: ATP; nucleotide binding; P-loop
F;163-373/Domain: FtsH/SEC18/CDC48-type ATP-binding domain homology <VATP>
F;190-197/Region: nucleotide-binding motif A (P-loop)

Query Match       78.7%;  Score 1665;  DB 2;  Length 406;
Best Local Similarity   78.5%;
Matches 328;  Conservative 38;  Mismatches 40;  Indels 12;  Gaps 2;

Qy     1 MALVGVELKHAAEGVPEANCSAKPTKQGSGLRHYSLNIHEHQLLLRQKTHNLNRLEAQR 60
Db     1 -------------------------------KAGSGLRQYLSKIEELQLIVNDKSQNLRLQAQR 49

Qy    61 NDLNSRVRMLREELQLLQEPGSYVGEVVKVKMGKNKVLVKVHPEGKYVVDIDKNIDITKIT 120
Db    50 NELNAKVRLLRRELQLLQEQGSYVGEVVRAMDKKKVLVKVHPEGKFVVDVDKNIDNDVT 109

Qy   121 PSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPI 180
Db   110 PNCRVALRNDSYTLHKILPNKVDPLVSLMMVEKVPDSTYEMIGGLDKQIKEIKEVIELPV 169

Qy   181 KHPELFEALGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKFIGEGAR 240
Db   170 KHPELFEALGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKFIGEGAR 229

Qy   241 MVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNK 300
Db   230 MVRELFVMAREHAPSIIFMDEIDSIGSSRLEGGSG--GDSEVQRTMLELLNQLDGFEATKN 288

Qy   301 IKVLMATNRIDILDQALLRPGRIDRKIEFPPTPNEESRLDILKIHSRRMNLMRGIDLKKIA 360
Db   289 IKVIMATNRIDMLDSALLRPGRIDRKIEFPPNEEARLDILKIHSRKMNLTRGINLRKIA 348

Qy   361 EKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
Db   349 ELMPGASGAEVKGVCTEAGMYALRERRVHVTQEDEEMAVAKVMQKDSEKNMSIKKLWK 406
```

```
RESULT 4
JN0610
probable transcription factor DdTBP10 - slime mold (Dictyostelium discoideum) (fragme
N;Alternate names: Tat-binding protein 10
C;Species: Dictyostelium discoideum
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Jan-2001
A;Accession: JN0610
R;Shaw, D.R.; Ennis, H.L.
 Biochem. Biophys. Res. Commun. 193, 1291-1296, 1993
A;Title: Molecular cloning and developmental regulation of Dictyostelium discoideum h
A;Reference number: JN0610; MUID:93312322; PMID:8323548
A;Accession: JN0610
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-389 <SHA>
A;Cross-references: GB:L16579; NID:g290056; PIDN:g290056; PID:AAA33254.1; PID:g290057
C;Superfamily: ATP-dependent 26S proteinase; FtsH/SEC18/CDC48-type ATP-binding domain
C;Keywords: ATP; nucleotide binding; nucleus; P-loop
A;Superfamily: ATP-dependent 26S proteinase; FtsH/SEC18/CDC48-type ATP-binding domain homology <VATP>
F;145-356/Domain: FtsH/SEC18/CDC48-type ATP-binding domain homology <VATP>
F;172-179/Region: nucleotide-binding motif A (P-loop)

Query Match       81.5%;  Score 1630;  DB 2;  Length 389;
Best Local Similarity   77.0%;
Matches 313;  Conservative 40;  Mismatches 31;  Indels 0;  Gaps 0;

Qy    35 YSLNIHEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEVVKVKMGKN 94
Db     6 YCSKIEELEIKVNERAQDLRRLEAQRNELNNRVRMLKEELQLLVNPEGSHVAAEVVKLMGKN 65

Qy    95 KVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKV 154
Db    66 KVLVKVNPEGKFVVDIDPTVDIAKLTPSTRAALKHESYTLHRILPNKIDPLVSLMKVEKI 125
```

```
Qy  155 PDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAV 214
        |||||||:|||||||||||||||||||||||||||:||||||||||||||||||||||||
Db  126 PDSTYDMVGGLDKQIKEIKEVIELPIKHPELFEALGIAQPKGVLLYGPPGTGKTLLARAV 185

Qy  215 AHHTDCTFIRVSGSELVQKIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGS 274
        |||||||||||||||||||:|||||||||||:||||||||||||||||||||:| |:||
Db  186 AHHTDCTFIRVSGSELVQKIIGEGSRMVRELFIWAREHAPSIIFMDEIDSIGSRGESGS 245

Qy  275 GNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPTNE 334
        | ||||||||||||||||||||:|||||:| ||||||||||||||||||||||||| :
Db  246 GGGDSEVQRTMLELLNQLDGFESTKNIKVMCTNRIDILDPALLRPGRIDRKIEFPNPGD 305

Qy  335 ESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHVTQED 394
        .|||||||||||:|||.|||:|||||:|||||||||||||||||||||||||||||||.||
Db  306 AGRLDILKIHSRKMNLTRGINLKKISDKMNGASGAELKAVCTEAGMYALRERRVHVSQED 365

Qy  395 FEMAVAKVMKKETEKNMSLRKLWK 418
        |||||.||||||:|:|||||||||
Db  366 FEMAVSKVMKKDSEQNMSINKLWK 389
```

RESULT 5
T43799
proteasome protein p45/SUG [imported] - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 02-Sep-2000
C;Accession: T43799
R;Kazahari, K.; Nomoto, K.; Ono, M.
Gene 198, 323-327, 1997
A;Title: Gene coding for the transcription factor, SUG/proteasome, p45 is located nearly
A;Reference number: 222687
A;Accession: T43799
A;Molecule type: DNA
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Residues: 1-374 <KAZ>
A;Cross-references: EMBL:AB000493; PIDN:BAA22935.1
A;Experimental source: strain Wistar, tissue brain
C;Genetics:
A;Note: intron positions not resolved (incomplete sequence)
C;Superfamily: ATP-dependent 26S proteinase; FtsH/SEC18/CDC48-type ATP-binding domain ho

Query Match   76.8%; Score 1624.5; DB 2; Length 374;
Best Local Similarity   83.7%; Pred. No. 2.1e-95;
Matches 314; Conservative 37; Mismatches 23; Indels 1; Gaps 1;

```
Qy   44 LLLRQKTHLNLRLEAQRNDLNSRVRMLREELQLLQEPGSYYGEVVKVWGKNKVLVKVHPE 103
        |||||  |:|  ||||||||||||||||||:||||||||| ||: ||:|||| ||||||
Db    1 LIVNDKSQNLRLQAQRNELNARVLREELQLLQEQGGSYVGEVVRAMDKKVLVKVHPE 60

Qy  104 GKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIG 163
        ||||:||:||||| |||||||||||||||||||| ||||||| |:|||||:||||: ||
Db   61 GKFVVDVDKNIDINDVTPNCRVALRNDSYTLHKILPNKVDPLVSLMVEKVPDSTYEMIG 120

Qy  164 GLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFI 223
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GLDKQIKEIKEVIELPVKHPELFEALGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFI 180

Qy  224 RVSGSELVQKIYGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGGNGDSEVQR 283
        |||||||| ||||||||||||||:||||||||||||||||||||:| ||||: |||||||:
Db  181 RVSGSELVQKFIGEGARMVRELFIWAREHAPSIIFMDEIDSIGSSRLEGGSG-GDSEVQR 239

Qy  284 TMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEPPTPNEESRLDILKI 343
        ||||||||||||||: ||||:||||||||||:|||||||||||||.|| |||||||||
Db  240 TMLELLNQLDGFEATKNIKVIMATNRIDILDSALLRPGRIDRKIEFPPNEEARLDILKI 299

Qy  344 HSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVM 403
        ||||:|||||||||||| |:|||||||||:||:||||||| || ||||||||||||||||
Db  300 HSRKMNLTRGINLRKIATELMPGASGAEVKGVCTEAGMYALRERRVHVTQEDFEMAVAKVM 359

Qy  404 KKETEKNMSLRKLWK 418
        ||||||||:||||||
Db  360 QKDSEKNMSIKKLWK 374
```

RESULT 6
S45176
26S proteinase regulatory subunit 8 homolog - fission yeast (Schizosaccharomyces pombe
N;Alternate names: transcription factor SUG1 homolog
C;Species: Schizosaccharomyces pombe
C;Date: 03-May-1994 #sequence_revision 02-Aug-1994 #text_change 19-Jan-2001
C;Accession: S45176; T39958
R;Michael, H.
submitted to the EMBL Data Library, September 1993
A;Reference number: S45176
A;Molecule type: DNA
A;Accession: S45176
A;Cross-references: EMBL:U02280; NID:g406050; PIDN:AAA61615.1; PID:g406051
R;Xiang, Z.; Aves, S.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, January 1999
A;Reference number: Z21893
A;Accession: T39958
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-403 <XIA>
A;Cross-references: EMBL:AL035065; PIDN:CAA22628.1; GSPDB:GN00067; SPDB:SPBC23G7.12c
A;Experimental source: strain 972h-; cosmid c23G7
C;Genetics:
A;Gene: SPBC23G7.12c
A;Introns: 18/3
A;Map position: 2
C;Superfamily: ATP-dependent 26S proteinase; FtsH/SEC18/CDC48-type ATP-binding domain
A;Keywords: ATP; nucleotide binding; P-loop
F;159-370/Domain: FtsH/SEC18/CDC48-type ATP-binding domain homology <VAT>
F;186-193/Region: nucleotide-binding motif A (P-loop)

Query Match   75.5%; Score 1597; DB 2; Length 403;
Best Local Similarity   79.4%; Pred. No. 1.3e-93;
Matches 312; Conservative 36; Mismatches 45; Indels 0; Gaps 0;

```
Qy   26 KQGEGLRHYYSLNIHEHQLLLRQKTHLNLRLEAQRNDLNSRVRMLREELQLLQEPGSYYG 85
        .: :.: :||.||: :| ||.|||:.|||||||||:||.||| | |||:||||||||| |
Db   11 QSNENIVQYYTQKIQDAELAILQKTQNLRLEAQRNGLNARVLRLLREEIQLLQEPGSYVG 70

Qy   86 EVVKVWGKNKVLVKVHPEGKYVVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPL 145
        ||:|  |:||||||||||||||||||   |||.:| |. |||||||||| | .||.||||
Db   71 EVIKTMGNKVLVKVHPEGKYVVVDISPDIDIKEIKPNIRVALRNDSYQLIKILPNKVDPL 130

Qy  146 VNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGT 205
        |.||.|||:|||||:|:|||::|||||||||||||:||||||:||||||||||||||||||
Db  131 VSLMMVEKIPDSTYEMVGGLEKQIKEIKEVIELPVKHPELFEALGIAQPKGVLLYGPPGT 190

Qy  206 GKTLLARAVAHHTDCTFIRVSGSELVQKIYGEGSRMVRELFVMAREHAPSIIFMDEIDSI 265
        |||||||||||||||.|||||||||||| |||||||||||||.|||||||||||||||||
Db  191 GKTLLARAVAHHTDCKFIRVSGSELVQKIIGEGSRMVRELFVWAREHAPSIIFMDEIDSI 250

Qy  266 GSARMESGSGGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDR 325
        ||:|.|| ||.|.||||||||||||||||||||: ||||:||||||||||:|||||||||
Db  251 GSSRSDSSGGSGDSEVQRTMLELLNQLDGFEATKNIKVIMATNRIDILDSALLRPGRIDR 310

Qy  326 KIEFPTPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRE 385
        |||||.|..::| :|:.||||:.||:||||||.|||||||||||||||:|||||||:||||
Db  311 KIEFPPPSAEARAEILRIHSRSMNLTRGIDLKSIAEKMNGASGAELKGVCTEAGMYALRE 370

Qy  386 RRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
        ||||||||||:|||||.: |.|.: |||:||.:
Db  371 RRVHVTQEDFELAVAKVLNKGDSGEMSLQKLFK 403
```

RESULT 7
S51042
tat-binding protein homolog - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 19-Jan-2001
C;Accession: S51042
R;Hirtzlin, J.; Faerber, P.M.; Franklin, R.M.

Eur. J. Biochem. 226, 673-680, 1994
A;Title: Isolation of a novel plasmodium falciparum gene encoding a protein homologous t
A;Reference number: S51042
A;Status: preliminary
A;Accession: S51042
A;Molecule type: mRNA
A;Residues: 1-435 <HIR>
A;Cross-references: EMBL:X77914
C;Superfamily: ATP-dependent 26S proteinase; FtsH/SEC18/CDC48-type ATP-binding domain ho
C;Keywords: ATP; nucleotide binding; P-loop
F;193-402/Domain: FtsH/SEC18/CDC48-type ATP-binding domain homology <VATP>
F;220-227/Region: nucleotide-binding motif A (P-loop)

Query Match   74.9%; Score 1584.5; DB 2; Length 435;
Best Local Similarity 72.6%; Pred. No. 8.6e-93;
Matches 312; Conservative 50; Mismatches 49; Indels 19; Gaps 3;

```
Qy      6 VELKHAABGVP------------------------------EANCSAKPTKQGE-GLRHYYSLNIHEHQLLLRQ 48
            :||:|||||||                              ||| |: :|||||||||||||||||||||||||
Db      8 VRSKNLSEGIKNEISEKQNMENNNNNNNDDNKNSSALNEEQVQSGIKRYYELKIEEYESIINK 67

Qy     49 KTHNLNRLEAQRNDLNSRVRMLREELQOLLQBPGSYVGEVAVKMGKNKVLVKVHPEGKYVV 108
           ||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||
Db     68 KLQNKRLEAQRNELNARVRELCDEIIQYLLEAASYVGEIVAKPMGKNKVLVKINPEGKYVV 127

Qy    109 DIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDDQQ 168
           :|| |: || ||||||||||||||||||||||||||:|||:|||||||:|||||||||||||
Db    128 DIARHINISHCTPNTRVALYNDSYKLHKILPSKVDPLVSLMKVEKVPDSTYEMVGGLDQQ 187

Qy    169 IKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGS 228
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    188 VKEVKEVIELPVKHPEIFESLGISOQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGS 247

Qy    229 ELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESSGSGNGDSEVQRTMLEL 288
           ||||||||||||||||||||||||||||||||||||||:||| ||  |||||||||||||
Db    248 ELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSQORIE--GEHGDSEVQRTMMEL 305

Qy    289 LNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEPPTPNEESRLDILKIHSRRM 348
           |||||||||||||:||||||||||||||||||||||||||| :||||||||:|||||||||
Db    306 LNQLDGFESTQNIKVIMCTNRIDILDEALLRPGRIDRKIEFPNPNVEARMEILKIHSRKI 365

Qy    349 NLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETE 408
           ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db    366 NLMRGIDMLKIATHMNNCSGAEVKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKQDAE 425

Qy    409 KNMSLRKLWK 418
           ||||||||||
Db    426 KNFTLRKLWK 435
```

RESULT 8
T27048
hypothetical protein Y49E10.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
A;Reference number: Z20303
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000
C;Accession: T27048
R;Barlow, K.
submitted to the EMBL Data Library, August 1997
A;Residues: 1-416 <WIL>
A;Cross-references: EMBL:Z98866; PIDN:CAB11558.1; GSPDB:GN00021; CESP:Y49E10.1
A;Experimental source: clone Y49E10
C;Genetics:
A;Gene: CESP:Y49E10.1
A;Map position: 3
A;Introns: 42/3; 300/3
C;Superfamily: ATP-dependent 26S proteinase; FtsH/SEC18/CDC48-type ATP-binding domain ho

Query Match   74.3%; Score 1572; DB 2; Length 416;
Best Local Similarity 75.7%; Pred. No. 5e-52;

Matches 306; Conservative 49; Mismatches 41; Indels 8; Gaps 2;

```
Qy     22 AKPTKQ------GEGLRHYYSLNIHEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLREEL 74
           :|||||       |||:||:|||||||||||||||||||||||||||||||||||||||
Db     14 SKPTAQKLTEESDEKTLRKYFSTKVDDAQOKVADRKSQONVRLOAQRNELNTKVRMLKEEL 73

Qy     75 QLLQEPGSYVGEVAVKMGKNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVL 134
           ||||||||||||||||:||||||||||||||||||||:|| ||| ||||||||||||||||
Db     74 QOLHEQGSYVGEVSKAMDKKKVLVKVHPEGKYVVDVDKSIDINSLNTGARVALRADSYAL 133

Qy    135 HLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDDQQIKEIKEVIELPIKHPELFESLGIAQP 194
           |||||||||||:||||||||||||||||||||||||:||||||||| |||||||:||||||
Db    134 HKLLPNKVDPLVSLMMVEKVPDSTYEMVGGLDKDIQKIKEIKEVIELPLKHPELFDALGIAQP 193

Qy    195 KGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAP 254
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    194 KGVLLFGPPGTGKTLLARAVAHHTECTFIRVSGSELVQKFIGEGARMVRELFVMAREHAP 254

Qy    255 SIIFMDEIDSIGSARMESSGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDIL 314
           |||||||||||||||  ||  |||||||||||||| |||||||||||||||:|||||||||
Db    254 SIIFMDEIDSIGSSRVE-GSSGGDSEVQRTMLELLNQLDGFEATKNIKVIMATNRIDILD 312

Qy    315 QALLRPGRIDRKIEPPTPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAV 374
           ||||||||||||||||| :||||||||:|||||||||||:|||:|:|||||:||:|||||
Db    313 PALLRPGRIDRKIEFPAPDEKARADILKIHSRKMNLMRGINMAKAIAEQIPGASGAEVKSV 372

Qy    375 CTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
           ||||||||||||||||||||||||||:||||||:|||||||||
Db    373 CTEAGMFALRERRIHVTQEDFEMAVGKVMQKDSEKNMSIKIKLWK 416
```

RESULT 9
T49402
probable 26S proteinase subunit (SUG1) protein (imported) - Neurospora crassa
N;Alternate names: protein B1D4.170
C;Species: Neurospora crassa
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jun-2000
C;Accession: T49402
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu
submitted to the Protein Sequence Database, May 2000
A;Reference number: Z25022
A;Accession: T49402
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-389 <SCH>
A;Cross-references: EMBL:AL355928; GSPDB:GN00116; NCSP:B1D4.170
A;Experimental source: BAC clone B1D4; strain OR74A
C;Genetics:
A;Gene: NCSP:B1D4.170
A;Map position: 6
A;Introns: 79/1
C;Superfamily: ATP-dependent 26S proteinase; FtsH/SEC18/CDC48-type ATP-binding domain

Query Match   72.2%; Score 1528.5; DB 2; Length 389;
Best Local Similarity 78.6%; Pred. No. 2.6e-89;

Matches 305; Conservative 30; Mismatches 52; Indels 1; Gaps 1;

```
Qy     31 LRHYYSLNIHEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEVAVKV 90
           |:|||:|||||||||:|||||||||||||||||||||||||||||||||||||||||||:|
Db      3 LDNYYHNKIEAMKLEILKGQAVLRRLEAQNDYNSRVRRLREELGLLQGQDIIKRNIDITKITPS 62

Qy     91 MGKNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMK 150
           |||||||||||||||||:| ||:||:|||||||||||||||:||||||||||||||:|||
Db     63 MSTKKVLVKVHPEGKYVVDSDNVDITKLTVGKRVTLLSDSYKLEKMLPSSVDPLVSLMM 122

Qy    151 VEKVPDSTYDMIGGLDDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLL 210
           |||||||||:|||||:|||||||||||||:||||||||||||:||||||||||||||||||
Db    123 VEKVPDSTYDMIGGLDQIKEIKEVIELGLKHPELFESLGIAQPKGVLLYGPPGTGKTLL 182

Qy    211 ARAVAHHTDCTFIRVSGSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARM 270
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183 ARAVAHHTDCKFIRVSGSSELVQKYIGEGSRMVREELGLKHPELFESLGIAQPKGVLLYGPPGTGKTLL 242
```

RESULT 10

S64052

26S proteasome regulatory particle chain RPT6 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein G3514; protein YGL048c; transcription factor SUG1
C;Species: Saccharomyces cerevisiae
A;Experimental source: strain S288C
R;Swaffield, J.C.; Bromberg, J.F.; Johnston, S.A.
Nature 357, 698-700, 1992
C;Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 19-Jan-2001
A;Title: Alterations in a yeast protein resembling HIV Tat-binding protein relieve requi...
R;Reference number: S24016; MUID:92310548; PMID:1614516
C;Accession: S64052; S24016; S28910; S35115
R;Feuermann, M.; Potier, S.; Souciet, J.L.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64044
A;Accession: S64052
A;Molecule type: DNA
A;Residues: 1-405 <FEU>
A;Cross-references: EMBL:Z72570; NID:g1322538; PIDN:CAA96750.1; PID:g1322539; MIPS:YGL04...
A;Accession: S24016
A;Molecule type: DNA
A;Residues: 1-405 <GOY>
A;Cross-references: EMBL:X66400; NID:g4590; PIDN:CAA47023.1; PID:g4591
A;Note: the authors did not translate the codons for residues 401-405
R;Swaffield, J.C.; Bromberg, J.F.; Johnston, S.A.
Nature 360, 768, 1992
A;Reference number: S28910; MUID:93101219; PMID:1465148
A;Contents: erratum
A;Accession: S28910
A;Molecule type: DNA
A;Residues: 401-405 <SW2>
R;Goyer, C.; Lee, H.S.; Malo, D.; Sonenberg, N.
DNA Cell Biol. 11, 579-585, 1992
A;Title: Isolation of a yeast gene encoding a protein homologous to the human Tat-bindin...
A;Reference number: S35115; MUID:93000478; PMID:1388730
A;Accession: S35115
A;Molecule type: DNA
A;Residues: 1-405 <GOY>
A;Cross-references: EMBL:L01626; NID:g172877; PIDN:AAA35138.1; PID:g172878
A;Cross-references: SGD:S0003016; MIPS:YGL048c
A;Map position: 7L
C;Superfamily: ATP-dependent 26S proteinase; FtsH/SEC18/CDC48-type ATP-binding domain ho...
C;Keywords: ATP; nucleotide binding; P-loop; transcription factor
F;162-372/Domain: FtsH/SEC18/CDC48-type ATP-binding domain homology <VATP>
F;187-202/Region: ATP binding #status predicted
F;189-196/Region: nucleotide-binding motif A (P-loop)

Query Match   71.1%;   Score 1504.5;   DB 2;   Length 405;
Best Local Similarity   74.3%;   Pred. No. 8.9e-88;
Matches   289;   Conservative   49;   Mismatches   50;   Indels   1;   Gaps   1;

```
Qy    271 ESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEPP 330
Db    243 E-GSSGGDSEVQRTMLELLNQLDGFEPTKNIKVIMATNRLDILDPALLRRGRIDRKIEPP 301

Qy    331 TPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERVHV 390
Db    302 PPSVEARADIIRHSRKMNLTRGINLTKIAEKMNGCSGAELKGVCTEAGMYALRERVHV 361

Qy    391 TQEDFEMAVAKVMKKETEKNMSLRKLWK 418
Db    362 TQEDFELATAKILNKHDDKEVSLAKLWR 389

Qy     30 GLRHYSLNHEHQLLRQKTHNLRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEVVK 89
Db     18 GIKPYFEQKIQETELJKIRSKFENVRLEAQRNALNDKVRIFKDELRLLQEPGSYVGEVIK 77

Qy     90 VMGKNKVLVKVHPEGKYVVYLDIKDNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLM 149
```

RESULT 11

T33633

hypothetical protein F56F11.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 31-Jan-2000
C;Accession: T33633
R;Latreille, P.; Kramer, J.; Keppler, D.
submitted to the EMBL Data Library, October 1998
A;Description: The sequence of C. elegans cosmid F56F11.
A;Reference number: 221379
A;Accession: T33633
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-443 <LAT>
A;Cross-references: EMBL:AF099922; PIDN:AAC71778.1; GSPDB:GN00021; CESP:F56F11.4
A;Experimental source: strain Bristol N2; clone F56F11
C;Genetics:
A;Gene: CESP:F56F11.4
A;Map position: 3
A;Introns: 69/3; 179/3; 267/3; 327/3
C;Superfamily: ATP-dependent 26S proteinase; FtsH/SEC18/CDC48-type ATP-binding domain

Query Match   70.1%;   Score 1482.5;   DB 2;   Length 443;
Best Local Similarity   77.1%;   Pred. No. 2.5e-86;
Matches   286;   Conservative   47;   Mismatches   37;   Indels   ,   1;   Gaps   1;

```
Qy     48 QKTHNLRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEVVKVKMGKNKVLVKVHPEGKYV 107
Db     74 EKSINVRRLMAQQKELNGKVRLLKEEISHLHEQGSHVGEVCKAMDRKKVLVKNHPDGKYI 133

Qy    108 VDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQ 167
Db    134 VDVAASVDISTLAAGTRVALRADNYAIHEVLESKVDPLVSLMVVEKVPDSTYEMVGGLDT 193

Qy    168 QIKIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSG 227
Db    194 QIKEIKEVIELPVKHPELFDALGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSG 253

Qy    228 SELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGGSGGNDSEVQRTMLE 287
Db    254 SELVQKFIGEGARMVRELFVMAREHAPSIIFMDEIDSIGSSRVE-GSRGGDSSEVQRTMLE 312

Qy    288 LLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPTPNEESRLDILKIHSRR 347
Db    313 LLNQLDGFEATKNIKVIMATNRLDILDSALLRPGRIDRKIEFPPAPDEKARAQILKIHSRK 372

Qy    348 MNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERERVHVTQEDFEMAVAKVMKKET 407
Db    373 MNLMRGIRMDKIAEQIPGASGAEVKSVCTEAGMFALRERRIHVTQEDFEMAVGKVMQKDS 432

Qy     78 IVSDDKVLVKVQPEGKYIVDVAKDINVKDLKASQRVCLRSDSYMLHKVLVSLM 137

Qy    150 KVEKVPDSTYDMIGGLDQQIKIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTL 209
Db    138 MVEKVPDSTYDMVGGLTRQIKIEKEVIELPVKHPELFELSLGIAQPKGVLLYGPPGTGKTL 197

Qy    210 LARAVAHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSAR 269
Db    198 LARAVAHHTDCKFIRVSGAELVQKYIGEGSRWVRELFVMAREHAPSIIFMDEIDSIGSTR 257

Qy    270 MESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEF 329
Db    258 VE-GSSGGDSEVQRTMLELLNQLDGFETSKNIKIIMATNRLDILDPALLRRGRIDRKIEF 316

Qy    330 PTPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERERVH 389
Db    317 PPPSVFAARAEILRIHSRKMNLTRGINLRKVAEKMNGCSGADVAGVCTEAGMYALRERIH 376

Qy    390 VTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
Db    377 VTQEDFELAVGKVMNKNQETAISVAKLFK 405
```

```
Qy        408 EKNMSLRKLWK 418
              |||||:|:|||||
Db        433 EKNMSIKKLWK 443
```

RESULT 12
F90131

```
Qy     26S proteasome SU [imported] - Guillardia theta nucleomorph
```

A;Species: nucleomorph Guillardia theta
C;Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
A;Note: The highly reduced genome of an enslaved algal nucleus.
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 15-Jun-2001
C;Accession: F90131
R;Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Rei
Nature 410, 1091-1096, 2001
A;Reference number: A99082; MUID:11323671; PMID:11323671
A;Accession: F90131
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-400 <DOU>
A;Cross-references: GB:AF083031; NID:g13794368; PIDN:AAK39745.1; GSPDB:GN00152
C;Superfamily: ATP-dependent 26S proteinase; FtsH/SEC18/CDC48-type ATP-binding domain ho
C;Keywords: nucleomorph
A;Gene: prs8
A;Genetics:
A;Genome: nucleomorph
A;Map position: 3

```
Query Match             50.9%;  Score 1076;  DB 2;  Length 400;
Best Local Similarity   59.1%;  Pred. No. 1e-60;
Matches  202;  Conservative  69;  Mismatches  71;  Indels  0;  Gaps  0;

Qy      61  NDLNSRVRMLREELQLLQEPGSYVGEVVKVMGKNKVLVKVKVHPEGKYYVVDIDKNIDITKIT  120
            ||||:: ::::::||||:||||||||| ::| | : ||||||  ||||||::|||:|||| :
Db      43  NVTNKKIIQFIKELSNLHKESSLLGEILKINSQNDIVKINPDGKFIILKLGKKINLKIK  102

Qy     121  PSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPVDSTYDMIGGLDQQIKEIKEVIELPI  180
            ||:||||||||:|||:|||||||||||||||||||||||||||||:|||||||||||||
Db     103  IGRDVAIRNESYEIHKVLPSKTDPLVSLMKIENVPFCNYNMIGGLESQIGKIKEIIELPL  162

Qy     181  KHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKYIGEGSR  240
            ||||||| ||||||||||||||||||||||||||||||||||||||||||||:||||:||
Db     163  KYPELFEILGVFQPKGILLYGPPGTGKTLIARAVAFHSNCSFIRVSGSELVQKYIGEGGR  222

Qy     241  MVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNK  300
            ||||||||||| |||:||||||||||||||||||||| :|||:|||||||||||||||||
Db     223  MVREIFSIAKKNSFSIIFMDEVDSIGSHRKKHVSSTGDSEVQRTMLELLNQLDQFEEHKN  282

Qy     301  IKVLMATNRIDILDQALLRPGRIDRKIEPPTPNEESRLDILKIHSRRMNLMRGIDLKKIA  360
            |||:|||||||||||:|||||||||||:|||||||||||:||:|| |||||||||||||
Db     283  IKILMATNRIDVLDPALIRPGRIDRKIPNPDVPEGRISILRIHLKLKICENGIDIWKIA  342

Qy     361  EKNMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKV  402
            |||:||:| |||||||||||||||||||:|||:||:||| |||
Db     343  KIIEGATGADIKAICTESGMSAIRKAKNIVSVEDILYAIDKI  384
```

RESULT 13
G71242

```
Qy    probable 26S proteinase regulatory subunit - Pyrococcus horikoshii
```

C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C;Accession: G71242
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: G71242
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-399 <KAW>
A;Cross-references: GB:AP000001; NID:g3236128; PIDN:BAA29270.1; PID:g3256587

A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBa
C;Genetics:
A;Gene: PH0201
C;Superfamily: ATP-dependent 26S proteinase; FtsH/SEC18/CDC48-type ATP-binding domain
F;154-363/Domain: FtsH/SEC18/CDC48-type ATP-binding domain homology <VAT>

```
Query Match             45.2%;  Score 955.5;  DB 2;  Length 399;
Best Local Similarity   52.6%;  Pred. No. 4.3e-53;
Matches  195;  Conservative  72;  Mismatches  95;  Indels  9;  Gaps  3;

Qy      46  LRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEVVKVKVMGKNKVLVKVHPEGK  105
            ||:|     :|  ||||||||||||||||:||||||||||||||||| ||||||||||||
Db      26  IRQELQVRMLEADKERLREELSRLRSEMSRLRQPPAFAGTVIEVLDEDRAIVQNYNGPR  85

Qy     106  YVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPVDSTYDMIGGL  165
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86  FVVRIAPWIDRKKLRPGTRVALDQRTMAVVEILPTSKDPTVLGFEVIERNVTYNDIGGL  145

Qy     166  DQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRV  225
            :|||||||||||||||||||||||||||||||||||||||||||||||:| |||||||||
Db     146  KKQLQELRAIELPLKHPELFEEVGIDPKGVLLYGPPGCGKTLMAKALAHEVNATFIRV  205

Qy     226  SGSSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTM  285
            |||||||||||||||||||||||||| :||:||||||||||:|:||||||  |||||||||
Db     206  VGSELVRKYIGEGARLVHELFELAKEKAPTIIFIDEIDAIGAKRMDETG -GEREVNRTL  264

Qy     286  LELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEPPTPNEESRLDILKIHS  345
            |||||||||||||||:|:||||||||||||:|||||||||||:|||||||||||:||:||
Db     265  MQLLAEMDGFDPRGNVKVIAATNRPDILDPALLRPGRFDRLIEVPLPDFEGRLEILKVHT  324

Qy     346  RRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKK  405
            ||||||||||||||||||||:| |||||||||||||||||||:|||:||:|||| |:||
Db     325  RRMKL-KGVDLRAIAEMTEGASGADLKAIATEAGMFAIRERRRTVVDKVLGN  383

Qy     406  ETEKNMSLRKL  416
            :::|||||||:|
Db     384  E--------RKL  387
```

RESULT 14
H75207

```
Qy    26S proteinase regulatory chain 4 PAB2233 - Pyrococcus abyssi (strain Orsay)
```

C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: H75207
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A;Reference number: A75001
A;Accession: H75207
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-399 <KAW>
A;Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49111.1; PID:g545
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: PAB2233
C;Superfamily: ATP-dependent 26S proteinase; FtsH/SEC18/CDC48-type ATP-binding domain
F;154-363/Domain: FtsH/SEC18/CDC48-type ATP-binding domain homology <VAT>

```
Query Match             44.3%;  Score 937.5;  DB 2;  Length 399;
Best Local Similarity   52.0%;  Pred. No. 5.9e-52;
Matches  193;  Conservative  71;  Mismatches  98;  Indels  9;  Gaps  3;

Qy      46  LRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEVVKVMGKNKVLVKVHPEGK  105
            ||:|      | ||||||||||||||||||:|||||||||||||||||||||||||:|| 
Db      26  IRQLELQVRMLEADKERLEREELSRLRSEMSRLRQPPAFAGTVIEVLDDDRAIVQNYNGPR  85

Qy     106  YVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPVDSTYDMIGGL  165
            |||||||||||||||||||||||||:||||:|||||||||||||||||||||||||||||
Db      86  FVVRIAPWIERDKLRPGARVALDQRTMAIIELLPSSKDPTVLGFEVIERPNVTYNDIGGL  145
```

```
Qy   166  DQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRV  225
          :|||:|||||||||||||||||||||||||||||||||||||: ||
Db   146  KKQLQEIREAIELPLLKHPELFEEVGIDPPKGVLLYGPPGCGKTLMAKALAHEVNATPIRV  205

Qy   226  SGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTM  285
          :|||||:|||||||:|:|:||: |:|:||:|||:||||:||:|:|:|:| : |:|:|:||
Db   206  VGSELVRKYIGEGARLVHELFELAKEKAPTIIFIDEIDAIGAKRNDETTG-GEREVNRTL  264

Qy   286  LELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPTPNESRLDILKIHS  345
          ::||:::|||:  :::||:|||||| ||||:|||||| ||:||::|:|: :||:|||::
Db   265  MQLLAEMDGFDPRGNVKVIAATNRPDILDPALLRPGRFDRLIEVPLPDFEGRLIEILKVHT  324

Qy   346  RRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKK  405
          |||:|  : DL::IAE  :GASGA::KA: TEAGMFA:RE:R  VTQEDF :AV KV: :
Db   325  RRMKL-KNVDLRVIAEITEGASGADIKAIATEAGMFAIREERTYVTQEDFLKAVDKVLGN  383

Qy   406  ETEKNMSLRKL  416
                 |||
Db   384  E------RKL  387
```

RESULT 15
G69496
ATP-dependent 26S proteinase regulatory subunit 4 homolog - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 19-Jan-2001
C;Accession: G69496
R;Klenk,H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: G69496
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-398 <KLE>
A;Cross-references: GB:AE000967; GB:AE000782; NID:g2689290; PIDN:AAB89280.1; PID:g264856
C;Superfamily: ATP-dependent 26S proteinase; FtsH/SEC18/CDC48-type ATP-binding domain ho
C;Keywords: ATP; nucleotide binding; P-loop
F;155-365/Domain: FtsH/SEC18/CDC48-type ATP-binding domain homology <VATP>
F;182-189/Region: nucleotide-binding motif A (P-loop)

Query Match   43.9%;   Score 929; DB 2; Length 398;
Best Local Similarity   51.3%;   Pred. No. 2e-51;
Matches 192; Conservative 64; Mismatches 112; Indels 6; Gaps 2;

```
Qy    34  YYSLNIHEHQLLRQKTHNLNRLEAQRNDLNSRVRMLRBELQLLQEPGSYVGEVKVMGK   93
          :|:|              :|         |:   |  || : :|   |  | :  |||
Db    20  YYKL------RELYRRLEDDEKKFIESERIRYERRVRLRSEVRLRSPPLLVGVVSDILED  74

Qy    94  NKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEK  153
          :|:||        |      :|  |   || :  | ||      |     |||    :||
Db    75  GRVVVKSSTGPKFVVVNTSQYINEEELKPGARVALNQQTLAIVNVLPTSKDPMVYGFEVEE  134

Qy   154  VPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARA  213
          :|:: ||:|||:|:|:::|::|:||| ||||||:::|| ||||||||||||||||||:||
Db   135  KPEVSYEDIGGLDVQIEEIREAVELPLLKPELFAEVGIEPPKGVLLYGPPGTGKTLLAKA  194

Qy   214  VAHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESG  273
          ||::| ||||||:|||:|||||||||::|||:|::|:|:||:|||:||:|:|:|:|| :
Db   195  VANQTRATFIRVGGSEFVQKYIGEGARLVREVFQLAKEKAPSIIFIDELDAIAARRTNSD  254

Qy   274  SGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPTPN  333
          :|: ::EVQRT ::LL::LDGF:  :: |V::ATNRIDILD:A:LRPGR DR:IE P P
Db   255  T-SGDREVQRTWMQLLAELDGFDPDRGDVKVIGATNRIDILDPAILRPGRFDRIIEVPLPT  313

Qy   334  EESRLDILKIHSRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHVTQE  393
          :|:|:|:: |||:|  ::|| | ::|: ::|||||::||||||||||||:|:    | |:
Db   314  FEGRIQIFKIHTRKMKLAEDVDPKELARITEGASGADIKAICTEAGMFAIREERAKVTML  373

Qy   394  DFEMAVAKVMKKET  407
          ||  |  || :||
Db   374  DFTKAIEKVLKKTT  387
```

OM protein - protein search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: November 9, 2003, 11:17:06 ; Search time 24 Seconds
(without alignments)
819.048 Million cell updates/sec

Title: US-09-462-972-2
Perfect score: 2116
Sequence: 1 MALVGVELKHAAEGVPEANC............VAKVMKKETEKNMSLRKLWK 418

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :   SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1680	79.4	406	1	PRS8_HUMAN	P47210 homo sapien
2	1660.5	78.5	405	1	PRS8_DROME	O18413 drosophila
3	1655.5	78.2	402	1	PRS8_MANSE	P54814 manduca sex
4	1644	77.7	461	1	PRS8_XENLA	P46470 xenopus lae
5	1630	77.0	389	1	PRS8_DICDI	P34124 dictyosteli
6	1611	76.1	414	1	PRS8_NAEFO	Q25544 naegleria f
7	1597	75.5	403	1	PRS8_SCHPO	P41836 schizosacch
8	1504.5	71.1	405	1	PRS8_YEAST	Q01939 saccharomyc
9	961.5	45.4	436	1	PSMR_METKA	Q8tx03 methanopyru
10	957.5	45.3	396	1	PSMR_PYRFU	Q8u4h3 pyrococcus
11	955	45.2	399	1	PSMR_PYRHO	O57940 pyrococcus
12	937.5	44.3	399	1	PSMR_PYRAB	Q9v287 pyrococcus
13	929	43.9	398	1	PSMR_ARCFU	O28303 archaeoglob
14	923	43.6	410	1	PSMR_METTH	O26824 methanobact
15	906	42.8	430	1	PSMR_METJA	O58576 methanococc
16	903	42.7	443	1	PRS4_CAEEL	O16368 caenorhabdi
17	901	42.6	440	1	PRS4_HUMAN	Q03527 homo sapien
18	897	42.4	440	1	PRS4_CHICK	Q90732 gallus gall
19	895	42.3	439	1	PRS4_DROME	P48601 drosophila
20	894.5	42.3	425	1	PRS7_PRUPE	O64982 prunus pers
21	894.5	42.3	426	1	PRS7_ARATH	Q9ssb5 arabidopsis
22	893	42.2	426	1	PRS7_ORYSA	Q9fxt9 oryza sativ
23	892	42.2	426	1	PRS7_SPIOL	O41365 spinacia ol
24	888	42.0	433	1	PRS7_HUMAN	P35998 homo sapien
25	888	42.0	433	1	PRS7_MOUSE	P46471 mus musculu
26	888	42.0	433	1	PRS7_XENLA	P46472 xenopus lae
27	881	41.6	433	1	PRS7_RAT	Q63347 rattus norv
28	880	41.6	439	1	PRSX_HUMAN	Q92524 homo sapien
29	878.5	41.5	435	1	PRSX_CAEEL	Q18787 caenorhabdi
30	864	40.8	406	1	PRSX_CAEEL	Q17071 caenorhabdi
31	859.5	40.8	406	1	PRSX_SCHPO	P36012 schizosacch
32	854	40.4	448	1	PRS4_ORYSA	P46466 oryza sativ
33	852	40.3	438	1	PRS7_SCHPO	O42931 schizosacch
34	850	40.2	420	1	PSMR_METMA	Q8pys8 methanosarc
35	850	40.2	421	1	PSMR_METAC	Q8t188 methanosarc
36	849.5	40.1	467	1	PRS7_YEAST	P33299 saccharomyc
37	848	40.1	437	1	PRS4_YEAST	P40327 saccharomyc
38	817	38.6	409	1	PSMR_AERPE	Q9yac7 aeropyrum p
39	816.5	38.6	437	1	SUG2_YEAST	P53549 saccharomyc
40	816.5	38.6	439	1	PRSA_HUMAN	P17980 homo sapien
41	816.5	38.6	439	1	PRSA_RAT	Q63569 rattus norv
42	811.5	38.4	429	1	PRSA_ORYSA	P46465 oryza sativ
43	809.5	38.3	404	1	PRSA_MOUSE	O88685 mus musculu
44	808.5	38.2	404	1	PRSB_XENLA	O42586 xenopus lae
45	807.5	38.2	392	1	PSMR_SULTO	Q97502 sulfolobus

ALIGNMENTS

RESULT 1
PRS8_HUMAN

PRS8_HUMAN   STANDARD;   PRT;   406 AA.
ID   PRS8_HUMAN
AC   P47210; O35051; P52915; P52916;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   26S protease regulatory subunit 8 (Proteasome subunit p45) (Thyroid
DE   hormone receptor interacting protein 1) (TRIP1) (MSUG1 protein)
DE   (TAT-binding protein homolog 10) (TBP10) (P45/SUG)
GN   PSMC5 OR SUG1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606; 10090, 10116, 9913, 9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=Human;
RX   MEDLINE=95246863; PubMed=7729537;
RA   Akiyama K., Yokota K., Kagawa S., Shimbara N., Demartino G.N.,
RA   Slaughter C.A., Noda C., Tanaka K.;
RT   "cDNA cloning of a new putative ATPase subunit p45 of the human 26S
RT   proteasome, a homolog of yeast transcriptional factor Sug1p.";
RL   FEBS Lett. 363:151-156(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=Human;
RX   MEDLINE=95295737; PubMed=7776974;
RA   Lee J.W., Choi H.-S., Gyuris J., Brent R., Moore D.D.;
RT   "Two classes of proteins dependent on either the presence or absence
RT   of thyroid hormone for interaction with the thyroid hormone
RT   receptor.";
RL   Mol. Endocrinol. 9:243-254(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=Human; TISSUE=Lung;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=Mouse;
RX MEDLINE=96176766; PubMed=8598193;
RA Vom Baur E., Zechel C., Heery D., Heine M., Garnier J., Vivat V.,
RA le Douarin B., Gronemeyer H., Chambon P., Losson R.;
RT "Differential ligand-dependent interactions between the AF-2
RT activating domain of nuclear receptors and the putative
RT transcriptional intermediary factors mSUG1 and TIF1.";
RL EMBO J. 15:110-124(1996).
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=Rat; TISSUE=Liver;
RX MEDLINE=96183075; PubMed=8607789;
RA Makino Y., Yogosawa S., Kanemaki M., Yoshida T., Yamano K.,
RA Kishimoto T., Moncollin V., Egly J.M., Muramatsu M., Tamura T.;
RT "Structures of the rat proteasomal ATPases: determination of highly
RT conserved structural motifs and rules for their spacing.";
RL Biochem. Biophys. Res. Commun. 220:1049-1054(1996).
RN [6]
RP SEQUENCE FROM N.A.
RC SPECIES=Rat; STRAIN=Wistar; TISSUE=Brain;
RX MEDLINE=98036065; PubMed=9370298;
RA Kazahari K., Nomoto K., Nakazato S., Ono M.;
RT "Gene coding for the transcription factor, SUG/proteasome, p45 is
RT located nearly 40 kb downstream from the rat growth hormone gene.";
RL Gene 198:323-327(1997).
RN [7]
RP SEQUENCE FROM N.A.
RC SPECIES=Bovine;
RX MEDLINE=98366204; PubMed=9701609;
RA Zhu X., Craft C.M.;
RT "Interaction of phosducin and phosducin isoforms with a 26S proteasome
RT subunit, SUG1.";
RL Mol. Vision 4:13-13(1998).
RN [8]
RP SEQUENCE FROM N.A.
RC SPECIES=Pig; TISSUE=Brain, and liver;
RX MEDLINE=96430086; PubMed=8833236;
RA Leeb T., Rettenberger G., Bruch J., Hameister H., Brenig B.;
RT "The porcine gene TBP10 encodes a protein homologous to the human
RT tat-binding protein/26S protease subunit family.";
RL Mamm. Genome 7:180-185(1996).

-!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
    DEGRADATION OF UBIQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
    COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
    26S COMPLEX.
-!- FUNCTION: PUTATIVE MEDIATOR OF TR-REGULATED TRANSCRIPTION. IT
    INTERACTS, IN VITRO, WITH THE THYROID HORMONE RECEPTOR (IN A
    THYROID HORMONE T3-DEPENDENT MANNER) AND WITH RETINOID X RECEPTOR
    (RXR).
-!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.

DR EMBL; AB000491; BAA22933.1; -.
DR EMBL; AB000493; BAA22935.1; -.
DR EMBL; AF069053; AAC19266.1; -.
DR EMBL; X97719; CAA61864.1; ALT_INIT.
DR PIR; S60343; S60343.
DR PIR; S61923; S61923.
DR PIR; S65536; S65536.
DR Genew; HGNC:9552; PSMC5.
DR MIM; 601681; -.
DR MGD; MGI:105047; Psmc5.
DR GO; GO:0005837; C:26S proteasome; TAS.
DR GO; GO:0008575; F:proteasome ATPase activity; TAS.
DR GO; GO:0003712; P:transcription cofactor activity; TAS.
DR GO; GO:0006366; P:transcription from Pol II promoter; TAS.
DR InterPro; IPR005937; 26S_p45.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_centr.
DR InterPro; IPR003960; AAA_sub.
DR Pfam; PF00004; AAA; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR01242; 26Sp45; 1.
DR PROSITE; PS00674; AAA; 1.
KW Proteasome; ATP-binding.
SQ SEQUENCE 406 AA; 45626 MW; 29C641CC4A85A7F7 CRC64;
FT NP_BIND 190 197 ATP (POTENTIAL).
FT CONFLICT 61 61 E -> R (IN REF. 1).
FT CONFLICT 197 197 E -> R (IN REF. 1).
FT CONFLICT 266 266 D -> S (IN REF. 2).
FT CONFLICT 272 272 T -> Q (IN REF. 2).
FT CONFLICT 300 300 I -> M (IN REF. 2).

```
Query                406 AA;  45626 MW;  29C641CC4A85A7F7 CRC64;

Query Match          79.4%;  Score 1680;  DB 1;  Length 406;
Best Local Similarity 79.2%;  Pred: No. 5.7e-100;
Matches  331;  Conservative   37;  Mismatches   38;  Indels   12;  Gaps    2;

Qy    1 MALVGVELKHAAEGVPEANCSAKPTKQGEGLRHYYSLNIHEHQLLLRQKTHNLNRLEAQR  60
          ||| |  :   || |            |:| | :||::| :|**:|**::|**:***|||
Db    1 MALDGPEQMELEEG------------KAGSGLRQYYLSKIEELQLIVNDKSQNLRLQAQR  49

Qy   61 NDLNSRVRMLREELQLLOEPGSYVGEVVKWNGKNKVLVKWHPEGKYVVDIDKNIDITKIT 120
         *:||::||:|******* |:|||||||:: ::||||||||||||:||| ||||*:*::*
Db   50 NELNAKVRLLREELQLLOEGGSYVGEVVRAMDKKKVLVKWHPEGKFVVDKNIDINDVT   109

Qy  121 PSTRVALRNDSYVLHLVLSKVDFLVNLMKVEKVPDSTYDMIGGLDDQIKEIKEVIELPI 180
         |  ||||||||.||  :|.*||.||.|| ||||||||:|||||||.||:||||||||| 
Db  110 PNCRVALRNDSYTLHKILDNKVDPLVSLMWVEKVPDSTYEMIGGLDKQIREIKEVIELPV 169

Qy  181 KHPELFESLGIAQPRGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKYIGEGSR 240
         |||||||:|||||||||||||||||||||||||||||||||||||||||||| |||| |
Db  170 KHPELFEALGIAQPRGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKFIGEGAR 229

Qy  241 MVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSSEVQRTMLELLNQLDGFEASNK 300
         |||||||||||||||||||||||||||:| ||| *||||*|||||||||||||||||*:*
Db  230 MVRELFVMAREHAPSIIFMDEIDSIGSSRLEGSG-GDSEVQRTMLELLNQLDGFEATKN   288

Qy  301 IKVLMATNRIDILDQALLRPGRIDRKIEFPTVNEESRLDILKIHSRRMNLMRGIDLKKIA 360
         |||:|||||||||.|||||||||||||||| *|||:||.|||||||:|||:||*:|:||
Db  289 IKVIMATNRIDILDSALLRPGRIDRKIEFPPNEEARLQILKIHSRKMNLTRGINLRKIA  348

Qy  361 EKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
         |:|.|||||:|.|||||||:|||||||||||||||||||||||:|::|||||:*|||
Db  349 ELMPGASGAEVKGVCTEAGMYALRERRVHVTQEDFEMAVAKVMQKDSEKNMSIKKLWK 406
```

DR EMBL; D44467; BAA07919.1; -.
DR EMBL; L38810; AAC41735.1; -.
DR EMBL; BC001932; AAH01932.1; -.
DR EMBL; BC002367; AAH02367.1; -.
DR EMBL; Z54219; CAA90061.1; -.
DR EMBL; Z54219; CAA90061.1; -.
DR EMBL; D83521; BAA11938.1; -.
DR EMBL; X89718; CAA61863.1; -.

RESULT 2
PRS8_DROME
ID   PRS8_DROME   STANDARD;   PRT;   405 AA.
AC   O18413; O45023;
DT   15-JUL-1998 (Rel. 36, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   26S protease regulatory subunit 8.
GN   PROS45 OR UG OR DUG OR CG1489.
OS   Drosophila melanogaster (Fruit fly).

OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=98137788; PubMed=9459929;
RA   Mounkes L.C., Fuller M.T.;
RT   "The DUG gene of Drosophila melanogaster encodes a structural and
RT   functional homolog of the S. cerevisiae SUG1 predicted ATPase
RT   associated with the 26S proteasome.";
RL   Gene 206:165-174(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Cheng L., Roemer N., Smyth K.-A., Belote J., Nambu J., Schwartz L.M.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley; TISSUE=Embryo;
RX   MEDLINE=22426066; PubMed=12537569;
RA   Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA   George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA   Rubin G.M., Celniker S.E.;
RT   "A Drosophila full-length cDNA resource.";
RL   Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
-!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
    DEGRADATION OF UBIQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
    COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
    26S COMPLEX (BY SIMILARITY).
-!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (Potential).

-!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

DR   EMBL; U97538; AAC48284.1; -.
DR   EMBL; AF043734; AAC63219.1; -.
DR   EMBL; AE003568; AAF50835.1; -.
DR   EMBL; AY051732; AAK93156.1; -.
DR   FlyBase; FBgn0020369; Pros45.
DR   GO; GO:0005838; C:19S proteasome regulatory particle; IDA.
DR   GO; GO:0004299; F:proteasome endopeptidase activity; IDA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IDA.
DR   InterPro; IPR005937; 26S p45.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR003960; AAA_sub.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   Proteasome; ATP-binding; Nuclear protein.
FT   NP_BIND   189   196   ATP (POTENTIAL).
FT   CONFLICT   94   94   F -> S (IN REF. 1).
SQ   SEQUENCE   405 AA; 45857 MW; C09EEE56B348211 CRC64;

Query Match   78.5%; Score 1660.5; DB 1; Length 405;
Best Local Similarity 82.4%; Pred. No. 9.9e-99;
Matches 323; Conservative 39; Mismatches 29; Indels 1; Gaps 1;

```
Qy   27  QGEGLRHYYSLNIHEHQLLRQKTHINLNRLEAQRNDLNSRVRMLREELQLLQEEGSYVGE   86
             |||:||||||||||||||||:|||||||||||| |||| ||||||||:|||||||||
Db   15  KGEGFRSYYIQKIEELQLVVAEKHQNLRRLQAQRNELNAKVRMLRBELQLLQBQGSYVGE   74

Qy   87  VVKVMGKNKVLVKVHPEGKVVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLV  146
         |||:| ||:||||:||||| |||:||||||  :|| ::|||||:||:||::||:|||||
Db   75  VVKPMDKKLKVLVKHPEGKEVVDLDKNIDINDVTPNCRVALRNESYTLHKILPNKVDPLV  134

Qy  147  NLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTG  206
          |||||||||||:|:||||||:||||||||||||:||||||::||||||||||||||||||
Db  135  SLMMVEKVPDSTYEMVGGLDKQIKEIKEVIELPVKHPELFDALGIAQPKGVLLYGPPGTG  194

Qy  207  KTLLARAVAHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIG  266
         |||||||||||:|||||||||||||||:||||||||||||||||||||||||||||||||
Db  195  KTLLARAVAHHTECTFIRVSGSELVQKFIGEGSRMVRELFVMAREHAPSIIFMDEIDSIG  254

Qy  267  SARMESGSGNQDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRK  326
         |:| | |:|  ||||||||||||||||||| :|||||:|||||||||| |||||||||||
Db  255  SSRIEGSSG-GDSEVQRTMLELLNQLDGFEATKNIKVIMATNRIDILDPALLRPGRIDRK  313

Qy  327  IEFPTPNEBSRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMPALRER  386
         ||||.|||:||||||||||||:|||:||:|:|||||:|||||||||:||||||||:||||
Db  314  IEFPPNBEARLDILKIHSRKMNLTRGINLRKIAELMPGASGAEVKGVCTEAGMYALRER  373

Qy  387  RVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK  418
         |||||||||||||||||::|.:|||||:|||||
Db  374  RVHVTQEDFEMAVAKVMQKDDSEKNMSIKKLWK  405
```

RESULT 3

PRS8_MANSE
ID   PRS8_MANSE   STANDARD;   PRT;   402 AA.
AC   P54874;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   26S protease regulatory subunit 8 (18-56 protein).
OS   Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).

Left column

```
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Sphingoidea;
OC   Sphingidae; Sphinginae; Manduca.
OX   NCBI_TaxID=7130;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96187870; PubMed=8606008;
RA   Sun D., Sathyanarayana U.G., Johnston S.A., Schwartz L.M.;
RT   "A member of the phylogenetically conserved CAD family of
RT   transcriptional regulators is dramatically up-regulated during the
RT   programmed cell death of skeletal muscle in the tobacco hawkmoth
RT   Manduca sexta.";
RL   Dev. Biol. 173:499-509(1996).
CC   -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC       DEGRADATION OF UBIQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC       COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC       26S COMPLEX (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (Potential).
CC   -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; U43728; AAC46996.1; -.
DR   InterPro; IPR005937; 26S_p45.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR003960; AAA_sub.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   Proteasome; ATP-binding; Nuclear protein.
FT   NP_BIND       186    193       ATP (POTENTIAL).
SQ   SEQUENCE   402 AA;  45338 MW;  4DE3F57E51FE235C CRC64;

Query Match             78.2%;  Score 1655.5;  DB 1;  Length 402;
Best Local Similarity   81.9%;  Pred. No. 2e-98;
Matches  321;  Conservative  41;  Mismatches  29;  Indels  1;  Gaps  1;

Qy     27 QGEGLRHYYSLNIHEHQLLRQKTHINLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVGE 86
          |||||||||:||||||||||:|||||||:|||:||:||:||:|||||||||||  |||||
Db     12 KGEGFRPYYITKIEELQLIVAEKSQNLRRLQAQRNELNAKRVRMLREELQLLQEQGSYVGE 71

Qy     87 VVKVMGKNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLV 146
          |||||:|:|||||||||||.|||:|:|:|:||.:.||||||||:||.|| ||| ::|||
Db     72 VVKPMDKKKVLVKVHPEGKEVVDLDNVDINDVTPNCRVALRNESYTLHKILPNKVDPLV 131

Qy    147 NLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTG 206
          .|||||||||||:|:|||||:||||||||||||:|||||||:|||||||||||||||||
Db    132 SLMMVEKVPDSTYEWVGGLDKQIKEIKEVIELPVKHPELFDALGIAQPKGVLLYGPPGTG 191

Qy    207 KTLLARAVAHHTDCTFIRVSGSSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIG 266
          ||||||||||||:||||||||||||||:|||||||||||||||||||||||||||||||:|
Db    192 KTLLARAVAHHTECTFIRVSGSSELVQKFIGEGSRMVRELFVMAREHAPSIIFMDEIDSIG 251

Qy    267 SARMESGSGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRK 326
          |:||||||:|:|:|:||||||||||||||||:|:|||:|:|||||||||||:||||||||||
Db    252 SSRIESGSG-GDSEVQRTMLELLNQLDGFEATKNIIVIMATNRIDILDPALLRPGRIDRK 310

Qy    327 IEFPPTNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRER 386
          ||||| :|||||||||:|||::|||:||:|:||||||:..|:.||:||||||||||:|||||
Db    311 IEFPPPNEEARLDILKIHSRKMNLTRGINLRKIAELMPGASGAEVKGVCTEAGMYALRER 386

Qy    387 RVHVTQEDFEMAVAKVMKKETEKNMSLRKLMK 418
          ||||||||||||||||:|:::::|:|:::||:
Db    371 RVHVTQEDFEMAVAKVMQKDSEKNNSIKKLMK 402
```

Right column

```
RESULT 4
ID   PR58_XENLA     STANDARD;      PRT;   461 AA.
AC   P46470; P79961;
DT   01-NOV-1995 (Rel. 32, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   26S protease regulatory subunit 8 (SUGI homolog) (xSUGI).
GN   SUG1.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary.
RA   Nacken W.K.F., Sorg C.;
RA   Nacken W.K.F., Kingsman A., Kingsman S., Sablitzky F., Sorg C.;
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   REVISIONS.
CC   -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC       DEGRADATION OF UBIQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC       COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC       26S COMPLEX (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (Potential).
CC   -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; X81986; CAA57512.1; -.
DR   InterPro; IPR005937; 26S_p45.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR003960; AAA_sub.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   Proteasome; ATP-binding; Nuclear protein.
FT   NP_BIND       185    192       ATP (POTENTIAL).
SQ   SEQUENCE   461 AA;  52248 MW;  5045D4EFFA4C64A2 CRC64;

Query Match             77.7%;  Score 1644;  DB 1;  Length 461;
Best Local Similarity   81.5%;  Pred. No. 1.3e-97;
Matches  321;  Conservative  39;  Mismatches  32;  Indels  2;  Gaps  2;

Qy     25 TKQGEGLRHYYSLNIHEHQLLRQKTHINLNRLEAQRNDLNSRVRMLREELQLLQEPGSYV 84
          :|||||||||||:||||||||||:||||||:|||:||:|||||||||||||||||  |||
Db      9 SRGGTGLRQYYLSKIEDLQLVNDKSQNLRRLQAQRNELNAKVRLLREELQLLQEQGSYV 68

Qy     85 GEVKVMGKNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDP 144
          |||:|:|:|:|||||||||||.|||:|:|:|:||.:.||||||||:|| || |:|:||
Db     69 GEVVRAMDKKKVLVKVHPEGKEVVDKNIDINDVTPNCRVALRNDSYTLHKILPNKVDP 128

Qy    145 LVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPG 204
          |.:|||||||||||:|:|||||:|||||||||||:||||||||:||||||||||||||||
Db    129 LVSLMMVEKVPDSTYEMIGGLDKQIKEIKEVIELPVKHPEHFEALGINQPKGVLLYGPPG 188

Qy    205 TGKTLLARAVAHHTDCTFIRVSGSSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDS 264
          ||||||||||||||:||||||||||||||:|||||||||||||||||||||||||||||:|
Db    189 TGKTLLARAVAHHTDCTFIRVSGSSELVQKFIGEGARMVRELFVMAREHAPSIIFMDEIDS 248

Qy    265 IGSARMESGSGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRID 324
          |||:||||||:|:|:|:||||||||||||||||:|:|||:||:|||||||||||:||||||
Db    189 IGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRID 324
```

RESULT 5
PRS8_DICDI

```
ID   PRS8_DICDI     STANDARD;      PRT;   389 AA.
AC   P34124;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   26S protease regulatory subunit 8 (TAT-binding protein homolog 10)
DE   (Fragment).
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RX   MEDLINE=93312222; PubMed=8323548;
RA   Shaw D.R., Ennis H.L.;
RT   "Molecular cloning and developmental regulation of Dictyostelium
RT   discoideum homologues of the human and yeast HIV1 Tat-binding
RT   protein.";
RL   Biochem. Biophys. Res. Commun. 193:1291-1296(1993).
CC   -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC       DEGRADATION OF UBIQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC       COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC       26S COMPLEX (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (Potential).
CC   -!- DEVELOPMENTAL STAGE: HIGHEST EXPRESSION IN VEGETATIVELY GROWING
CC       CELLS. THE LEVEL OF EXPRESSION FALLS STEADILY THROUGHOUT
CC       MULTICELLULAR DEVELOPMENT AND ARE NOT FOUND IN DORMANT OR
CC       GERMINATING SPORES.
CC   -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
```

```
DR   EMBL; L16579; AAA33254.1; -.
DR   PIR; JN0610; JN0610.
DR   Dictydb; DD01053; tbpC.
DR   InterPro; IPR005937; 26S_p45.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR003960; AAA_sub.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   Proteasome; ATP-binding; Nuclear protein.
FT   NON_TER       1      1
FT   NP_BIND     172    179       ATP (POTENTIAL).
SQ   SEQUENCE   389 AA;  43556 MW;  19D27D78EAF60C18 CRC64;

Query Match             77.0%;
Best Local Similarity   81.5%;
Matches  313;  Conservative  40;  Mismatches  31;  Indels   0;  Gaps   0;

                        Score 1630;  DB 1;  Length 389;
                        Pred. No. 8.2e-97;

Qy  249 IGS-RLEGGSG-GDSEVQRTMLELLNQLDGFEATKNIKVIMATNRIDILDSALLRPGRID 306
        ||| |:|.||| |||||||||||||||||:||||||||.|||||||||||||||.|||||
Db  325 RKIEFPTPNEESRLDILKIHSRRNNLMRGIDLKKIAEKNNGASGAELKAVCTEAGMFALR 384

Qy  307 RKIEFPPNEEARLDILKIHSRKMNLTRGINLRKIAELMPGASGAEVKGVCTEAGMYALR 366
        |||||.|||:|||||||||||:|||:||||||:|||:|||||||:||.||||||||:||
Db  385 ERRVHVTQEDFEMAVAKVMKKSETEKNMSLRKLWK 418

Qy  367 ERRVHVTQEDFEMAVAKVMKQKDSEKNMSIKKLWK 400
        |||||||||||||||||||||:||:||||||:||||
Db  385 ERRVHVTQEDFEMAVAKVMKKSETEKNMSLRKLWK 418
```

RESULT 6
PRS8_NAEFO

```
ID   PRS8_NAEFO     STANDARD;      PRT;   414 AA.
AC   Q25544;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   26S protease regulatory subunit 8 homolog (TAT-binding protein
DE   homolog).
OS   Naegleria fowleri.
OC   Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria.
OX   NCBI_TaxID=5763;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LEE mp;
RA   Sullivan P.K., Shaw D.R., Marciano-Cabral F., Ennis H.L.;
RT   "Isolation and characterization of a Naegleria fowleri Tat-binding
RT   protein cDNA.";
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC       DEGRADATION OF UBIQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC       COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC       26S COMPLEX (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (Potential).
CC   -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
```

```
DR   EMBL; U41812; AAB01762.1; -.
DR   InterPro; IPR005937; 26S_p45.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR003960; AAA_sub.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   Proteasome; ATP-binding; Nuclear protein.

Qy   35 YSINIHEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLREBLQLLLQEPDGGSYVGEVVKVMGKN 94
Db    6 YCSKIEBLEIKVNEKAQDLRRLEAQRNELNNRVRMLKEEBLQLLITNPGSHVABVVKLMGKN 65

Qy   95 KVLVKVHPEGKYVVDIDXNIDITKLTPSTRVALRNDSYVLHLVLPSKVDFLVNLMKVEKV 154
Db   66 KVLVKVNPEGKFVVDIDPTVDIAKLTPSTRAALKHESYTLHRILPNKIDFLVSLMKVEKI 125

Qy  155 PDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAV 214
Db  126 PDSTYDMVGGLDKQIKEIKEVIELPIKHPELFEBSLGIAQPRGVLLYGPPGTGKTLLARAV 185

Qy  215 AHHTDCTFIRVSGSELVQKYIGBGSRMVRELFVMAREHAPSIIFWDEIDSIGSARMESGS 274
Db  186 AHHTDCTFIRVSGSELVQKYIGEGSRNVRELPIMAREHAPSIIFMDEIDSIGSSRGESGS 245

Qy  275 GNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPTPNE 334
Db  246 GGGDSEVQRTMLELLNQLDGFESTKNIKVLMCTNRIDILDPALLRPGRIDRKIEFPNPGD 305

Qy  335 ESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHVTQED 394
Db  306 AGRLDILKIHSRKMNLTRGINLKKISDKMNGASGAELKAVCTEAGMYALRERRVHVSQED 365

Qy  395 FEMAVAKVMKKSETEKNMSLRKLWK 418
Db  366 FEMAVSKVMKKDSEQNMSINKLWK 389
```

```
FT      NP_BIND          197    204      ATP (POTENTIAL).
SQ      SEQUENCE    414 AA;  45999 MW;  628D3A2D3B3024C3 CRC64;
        Best Local Similarity   76.1%;   Score 1611; DB 1;  Length 414;
        Matches 309; Conservative  39; Mismatches  41; Indels   0;  Gaps   0;

Qy      30  GLRHYYSLNIHEHQLLLROKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEEVVK  89
                :||:||||||||||||||||||||| |||||:|||||||||||||||||||||||||:|||
Db      26  GLKNYIIAKIQEKEIVLRDKIENLRLQAQRNELNGRVRSLKEELQQLQEPGSHVGEEVVK  85

Qy      90  VMGKNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLM  149
                |||||||||||||||||:||:|||||| ||:| |:||||||||||:||||||||||||||
Db      86  VMGKKKVLIKISQEGKYIVDLDKSIELKDIITPGVRVALRNDSYALHKILPNKIDPLVSLM  145

Qy      150 KVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTL  209
                |||||||||||:|||||||||||||||||||||||||||:||||||||||||||||||:|
Db      146 KVEKVPDATYEMVGGLDLEQIKEIKEVIELPIKHPELFESLGIAQPKGVIMYGPPGTGKTL  205

Qy      210 LARAVAHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSAR  269
                |||||||||||||||||||||||||||||||||||||||||||:|||||||||||||:|
Db      206 LARAVAHHTDCTFIRVSGSELVQKYIGEGSKWVRELFVMAREHAPSIIFMDEIDSIGSTR  265

Qy      330 PTPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGABLKAVCTEAGMFALRERRVH  389
                |||:|||||||||:||||||||||||||||| |:|||:||||||||||||||||||||||
Db      326 PNPNELARIDILKIHSRRMNLTRGINLRKIASMLGASAGABIKATCTEAGMFALRERRVH  385

Qy      390 VTQEDFEMAVAKVMKKETEKNMSLRKLWK  418
                ||||:|||:||||||||:|:||:|:||:|
Db      386 VTQEDLEMAVAKVMRGDESNMSFKKLFK  414

RESULT 7
PR58_SCHPO

ID      PR58_SCHPO     STANDARD;   PRT;   403 AA.
AC      P41836;
DT      01-NOV-1995 (Rel. 32, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      26S protease regulatory subunit 8 homolog (Protein let1).
GN      LET1 OR SPBC23G7.12C.
OS      Schizosaccharomyces pombe (Fission yeast).
OC      Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC      Schizosaccharomycetales; Schizosaccharomycetaceae;
OC      Schizosaccharomyces.
OX      NCBI_TaxID=4896;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=972;
RX      MEDLINE=21848401; PubMed=11859360;
RA      Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA      Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA      Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA      Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA      Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA      Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA      James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA      Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA      Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA      Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA      Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA      Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA      Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA      Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA      Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA      Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA      Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA      Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA      Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA      Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA      Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA      Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA      Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA      Shpakovski G.V., Ussery D., Barrel B.G., Nurse P.;
RL      "The genome sequence of Schizosaccharomyces pombe.";
RL      Nature 415:871-880(2002).
CC       -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC           DEGRADATION OF UBIQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC           COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC           26S COMPLEX (BY SIMILARITY).
CC       -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (potential).
CC       -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC     ---------------------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC     ---------------------------------------------------------------------------
DR      EMBL; U02280; AAA61615.1; -.
DR      EMBL; AL035065; CAA22628.1; -.
DR      PIR; S45176; S45176.
DR      GeneDB_SPombe; SPBC23G7.12c; -.
DR      InterPro; IPR005937; 26S_p45.
DR      InterPro; IPR003593; AAA_ATPase.
DR      InterPro; IPR003959; AAA_ATPase_centr.
DR      InterPro; IPR003960; AAA_sub.
DR      Pfam; PF00004; AAA; 1.
DR      SMART; SM00382; AAA; 1.
DR      TIGRFAMs; TIGR01242; 26Sp45; 1.
DR      PROSITE; PS00674; AAA; 1.
KW      Proteasome; ATP-binding; Nuclear protein.
FT      NP_BIND         186    193      ATP (POTENTIAL).
SQ      SEQUENCE    403 AA;  45069 MW;  36DEF12C39841315 CRC64;

        Best Local Similarity   75.5%;   Score 1597; DB 1;  Length 403;
        Matches 312; Conservative  36; Mismatches  45; Indels   0;  Gaps   0;

Qy      26  KQGEGLRHYYSLNIHEHQLLLROKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVG  85
                :||:||||||||||||||||||| |||||||:|||||||||||||||||||||||||||
Db      11  QSNENIVQYYTQKIQODAELAILQKTQNLRLEAQRNGLNARVRLLREELQLLQEPGSYVG  70

Qy      86  EVVRKVMGKNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPL  145
                |:|| ||||||||||||||||||||||:||||:|||:|:||||||||||||||||||:||
Db      71  EVIKTMGKNKVLVKVHPEGKYVVDISPDIDIKEIKPNIRVALRNDSYOLIKILPNKVDPL  130

Qy      146 VNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGT  205
                |:|:|||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db      131 VSLMWEKIPDSTYEMVGGLEKQIKEIKEVIELPVKHPELFESLGIPQPKGILLYGPPGT  190

Qy      206 GKTLLARAVAHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSI  265
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      191 GKTLLARAVAHHTDCKFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSI  250

Qy      266 GSARMESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATWNRIDILDQALLRPGRID  325
                ||:|:||||||||||||||||||||||||||||||||||:||||||||||||||||||||
Db      251 GSSRSDSSGGSDSEVQRTMLELLNQLDGPEATNIKVIMATRNRIDILDPALLRPGRIDR  310

Qy      326 KIEEPTNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGABLKAVCTEAGMFALRE  385
                |||| |:||||||||||||||||| |||
```

```
RESULT 8

Qy          311 KIEFPPPSAEARAEILRIHSRSMNLTRGIDLKSIAEKMNGASGAELKGVCTEAGMFALRE 370
                ||||||:||:| ||||:|||||||||||||:||||||||:||:||||||||||||
Db          371 RRVHVTQEDPELAVAKVLNKGDSGEMSLQKLFK 403

Db          311 KIEFPPPSAEARAEILRIHSRSMNLTRGIDLKSIAEKMNGASGAELKGVCTEAGMFALRE 370

Qy          386 RRVHVTQEDPEMAVYAKVMKKETENXMSLRKLWK 418
                |||||||||||:|:||||||||||:|:|||:|:|
Db          371 RRVHVTQEDPELAVAKVLNKGDSGEMSLQKLFK 403
```

PR88_YEAST
ID   PR88_YEAST   STANDARD;   PRT;   405 AA.
AC   Q01939;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   26S protease regulatory subunit 8 homolog (SUG1 protein) (CIM3
DE   protein) (TAT-binding protein TBY1).
GN   FP16 OR SUG1 OR TBY1 OR CIM3 OR CRL3 OR YGL048C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92310548; PubMed=1614516;
RA   Swaffield J.C., Bromberg J.F., Johnston S.A.;
RT   "Alterations in a yeast protein resembling HIV Tat-binding protein
RT   relieve requirement for an acidic activation domain in GAL4.";
RL   Nature 357:698-700(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93000478; PubMed=1388730;
RA   Goyer C., Lee H.S., Malo D., Sonenberg N.;
RT   "Isolation of a yeast gene encoding a protein homologous to the human
RT   Tat-binding protein TBP-1.";
RL   DNA Cell Biol. 11:579-585(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=97377993; PubMed=9334674;
RA   Feuermann M., de Montigny J., Potier S., Souciet J.-L.;
RT   "The characterization of two new clusters of duplicated genes
RT   suggests a 'Lego' organization of the yeast Saccharomyces cerevisiae
RT   chromosomes.";
RL   Yeast 13:861-869(1997).
RN   [4]
RP   CHARACTERIZATION.
RA   Gerlinger U.-M., Wolf D.H., Hilt W.;
RT   "The proteasome is linked to cycloheximide resistance in yeast: CRL3
RT   is a subunit of the 26S proteasome.";
RL   Enzyme Protein 48:317-317(1995).
RN   [5]
RP   FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
RA   Rubin D.M., Coux O., Wefes I., Hengartner C., Young R.A.,
RA   Goldberg A.L., Finley D.;
RT   "Identification of the gal4 suppressor Sug1 as a subunit of the yeast
RT   26S proteasome.";
RL   Nature 379:655-657(1996).
CC   -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC   DEGRADATION OF UBIQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC   COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC   26S COMPLEX (BY SIMILARITY).
CC   -!- SUBUNIT: MAY FORM A HOMODIMER OR A HETERODIMER WITH A RELATED
CC   FAMILY MEMBER.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (Potential).
CC   -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --

DR   EMBL; X66400; CAA47023.1; -.
DR   EMBL; L01626; AAA35138.1; -.
DR   EMBL; Z72570; CAA96750.1; -.
DR   PIR; S64052; S64052.
DR   SGD; S0003016; RP T6.
DR   GO; GO:0008540; C:19S proteasome regulatory particle, base su...; IDA.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   InterPro; IPR005937; 26S_p45.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR003960; AAA_sub.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   Proteasome; Atp-binding; Nuclear protein.
FT   NP_BIND   189   196   ATP.
FT   CONFLICT   41   41   V -> G (IN REF. 1).
SQ   SEQUENCE   405 AA; 45271 MW; 9C59E92A0794F60F CRC64;

Query Match   71.1%; Score 1504.5; DB 1; Length 405;
Best Local Similarity   74.3%; Pred. No. 8.4e-89;
Matches 289; Conservative 49; Mismatches 50; Indels 1; Gaps 1;

```
Qy           30 GLRHYSLNIHEHQLLLROKTHNLRLEAQRNDLNSRVRNLREELQLLQEPGSYVGEVVK 89
                |:|:||||  |||:  |||||:|| |:::|||| :|:|||||||:::||:|:|||||
Db           18 GIKPYPEQKIQETELKIRSKTENVRRLEAQRNALNDKVRFIKDELRLLQEPGSVVGEVTK 77

Qy           90 VMGKNKVLVKVHPEGKYVVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLM 149
                |:  ||||| |:||||:|||||||:|||| |:|| |||||:|| ::|:||||:|||||:|
Db           78 IVSDKKVLVKVQPEGKYIVDVAKDINVKDLKASGQRVCLRSDSYMLHKVLENKADPLVSLM 137

Qy          150 KVEKVPDST*YDMIGGLDQQIKEIVKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTL 209
                |||||||||| :||:|||| |||||||:||:|| ||||||||||:||||||||||||||||
Db          138 MVEKVPDSTYDMVGGLTKQIKEIKEVIELPVKHPELFESLGIAQPKGVILYGPPGTGKTL 197

Qy          210 LARAIAHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSAR 269
                ||||:||||| ||||| |||||| |||:||:|||||||||:|||||||:||||:|| :|
Db          198 LARAVAHHTDCKFIRVSGAELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSTR 257

Qy          270 MESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEF 329
                ||| |:|||||||||||||||||||| |::|||:||||||||||||:|||||||||||||
Db          258 VE-GSGGGDSEVQRTMLELLNQLDGFETSKNIKIIMATNRLDILDPALLRPGRIDRKIEF 316

Qy          330 PTPNEESRLDILKIHSRRMNLMRGIDLKIIAEKMNGASGGAELKAVCTEAGMFALRERVH 389
                || |:||| |||:||||:|| |||:||| |||||||||| ||||||||||||||||||:|
Db          317 PPPSVAARAEILRIHSRKMNLTRGINLRKVAEKMNGCCSGADVKGVCTEAGMYALRERRH 376

Qy          390 VTQEDPEMAVYAKVMKKETENXMSLRKLWK 418
                ||||||:||||||:|||:||| ||||||:|
Db          377 VTQEDPELAVGKKVMNKQEFAISVAKLFK 405
```

RESULT 9

PSMR_METKA
ID   PSMR_METKA   STANDARD;   PRT;   436 AA.
AC   Q8TX03;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Proteasome-activating nucleotidase (Proteasome regulatory subunit).
GN   PAN OR MK0878.
OS   Methanopyrus kandleri.
OC   Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC   Methanopyrus.
OX   NCBI_TaxID=2320;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AV19 / DSM 6324 / JCM 9639;
RX   MEDLINE=21927647; PubMed=11930014;
RA   Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,

```
RA   Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA   Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA   Malykh A.G., Koonin E.V., Kozyavkin S.A.;
RT   "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT   and monophyly of archaeal methanogens.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
CC   -!- FUNCTION: Required for the ATP- or CTP-dependent degradation of
CC       proteins, but not small peptides, by the 20S proteasome (By
CC       similarity).
CC   -!- SUBUNIT: Homohexamer (Potential).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE010377; AAM02091.1; -.
DR   HAMAP; MF_00553; -; 1.
DR   InterPro; IPR005937; 26S_p45.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR003960; AAA_sub.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   Proteasome; ATP-binding; Complete proteome.
FT   NP_BIND     220    227    ATP (POTENTIAL).
SQ   SEQUENCE   436 AA;  49767 MW;  AE150086I6A02ED CRC64;

Query Match             45.4%;  Score 961.5;  DB 1;  Length 436;
Best Local Similarity   49.3%;  Pred. No. 3.3e-54;
Matches 202;  Conservative 77;  Mismatches 108;  Indels 23;  Gaps 6;

Qy    29  EGLRHYY-----------SININEHQLLL---RQKTHNLNRLEAQRN---DLNSRVR------ 68
Db    28  ELIKEYFKRLEELERKLRAHEERKLRIEARRRKTLE-KELEMBRDEKAELREELRRKEVMI 86

Qy    69  -MLREELQLLQEPGSYVGESVVKWMGKNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVAL 127
Db    87  EKLRSDLQRMKKPLIVGTVEEIIDDGRVIVKSSTGPKFVSNVSPTVDRNELEPGANVAL 146

Qy   128  RNDSYVLHLVLPSKVDPLVNLMKVEKPVPDSTYDMIGGLDQQIKEIKEVIELEIKHPELFE 187
Db   147  NQQSMAVDVLPSEKDSRVLAMEVDESPDVSYDGIGGLDEQIREIREVEKPLKEPELFE 206

Qy   188  SLGIAQPRGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKIYGEGGSRMVRELFV 247
Db   207  KVGVEPPRGVLLYGPPGTGKTLLAKAVANHADATFIRLAAPELVQKFIGEGARLVRELFE 266

Qy   248  MAREHAPSIIIFMDEIDSIGSARMESGSGGNGDSEVQRTMLELLNQLDGFEASNNIKVLMAT 307
Db   267  LAREKAPSIIFIDEIDAIGARRMRDAT-SGDRRVQRITLQLLAEMDGFPLDDIKVIAAT 325

Qy   308  NRIDILDQALLRPGRIDRKIEFPPTPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGAS 367
Db   326  NRKDILDPALLRPGRFDRHIKIPLPDEEGRYEIFKIHTRDMNLAEDVDLQLKAKITEGAS 385

Qy   368  GAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLW 417
Db   386  GADIKAICTEAGMAIREDRDIVTMDDFLKAVDRVMGKKEEESGEFKRAY 435
```

```
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Proteasome-activating nucleotidase (Proteasome regulatory subunit).
GN   PAN OR PF0115.
OS   Pyrococcus furiosus.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=2261;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA   Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT   "The complete sequence of the Pyrococcus furiosus genome.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Required for the ATP- or CTP-dependent degradation of
CC       proteins, but not small peptides, by the 20S proteasome (By
CC       similarity).
CC   -!- SUBUNIT: Homohexamer (Potential).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE010137; AAL80239.1; -.
DR   HAMAP; MF_00553; -; 1.
DR   InterPro; IPR005937; 26S_p45.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR003960; AAA_sub.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   Proteasome; ATP-binding; Complete proteome.
FT   NP_BIND     178    185    ATP (POTENTIAL).
SQ   SEQUENCE   396 AA;  44804 MW;  791B3666AB853F96 CRC64;

Query Match             45.3%;  Score 957.5;  DB 1;  Length 396;
Best Local Similarity   52.4%;  Pred. No. 5.2e-54;
Matches 194;  Conservative 75;  Mismatches 96;  Indels 5;  Gaps 3;

Qy    46  LRQKTHNLNRLEAQRNDLNSRVRMLRELQLLQEPGSYVGEVVKMGKNKVLVKVHPEGK 105
Db    23  IRQLELQVRMLEADKERLERELSRLRSEMSRLRQPPAFAGSVIEVLDDDRAIVQNYNGPR 82

Qy   106  YVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGL 165
Db    83  FVVRIAPWIDKSKLRPGTRVALDQRTMAIIEILPASKDPAVLGFEVVERPNVTYNDIGGL 142

Qy   166  DQQIKEIKEVIELPIKHPELFESSLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRV 225
Db   143  KKQLQELREAIELPLKHPELFEEVGIDPPKGVLLYGPPGCGKTLMAKAIAHHVNATFIRV 202

Qy   226  SGSSELVQKIYGEGGSRMVRELFVWAREHAPSIIIFMDEIDSIGSARMESGSGNGDSEVQRTM 285
Db   203  VSSELVRKYIGEGARLVHELFELAKERXAPTIIFIDEIDAIGAKRLDETTG-GEREVRTL 261

Qy   286  LELLNQLDGFEASNNIKVLMATNRIDILDQALLRPGRIDRKIEFPPTPNEESRLDILKIHS 345
Db   262  MQLLAEMDGFDPRGNVKVIAATNRPDILDPALLRPGRFDRLIEVPLPDFEGRLEILKVHT 321

Qy   346  RRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVM- 403
Db   322  RRMKL-RGVDLRLIAELTEGASGADLKAIATEAGMFAIRERRTYVTQEDFLKAIDKVLGN 380

Qy   404  -KKETEKNMS 412
```

```
RESULT 10
ID   PSMR_PYRFU     STANDARD;     PRT;   396 AA.
AC   Q8U4H3;
DT   28-FEB-2003 (Rel. 41, Created)
PSMR_PYRFU
```

RESULT 11

```
Db     381  EKKIIQQIMS 390

ID   PSMR_PYRHO     STANDARD;      PRT;   399 AA.
AC   O57940;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Proteasome-activating nucleotidase (Proteasome regulatory subunit).
GN   PAN OR PH0201.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=53953;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE=98344137; PubMed=9679194;
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
CC   -!- FUNCTION: Required for the ATP- or CTP-dependent degradation of
CC       proteins, but not small peptides, by the 20S proteasome (By
CC       similarity).
CC   -!- SUBUNIT: Homohexamer (Potential).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
```

```
DR   EMBL; AP000001; BAA29270.1; -.
DR   PIR; G71242; G71242.
DR   HAMAP; MF_00553; -; 1.
DR   InterPro; IPR005937; 26S_p45.
DR   InterPro; IPR005593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR003960; AAA_sub.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   Proteasome; ATP-binding; Complete proteome.
FT   NP_BIND     181    188       ATP (POTENTIAL).
SQ   SEQUENCE   399 AA;  45177 MW;  5D2F62F7F530C42E CRC64;

Query Match    45.2%;  Score 955.5;  DB 1;  Length 399;
Best Local Similarity   52.6%;  Pred. No. 7e-54;
Matches  195;  Conservative   72;  Mismatches   95;  Indels    9;  Gaps    3;

Qy    46  LRQKTHLNLRLEAQRNDLNSRVRMLRREELQLLQEPGSSYVGEVKVWGKNKVLVKVHPEGK 105
Db    26  IRQLELQVRMLEADKERLEREISRLRSEMSRLRQPPAFAGTVIEVLDEDRAIVQNYNGPR  85

Qy   106  YVVDIDKNIDITKITFSTRVALRNDSYVLHVLPSKVDPLVNLMKVEKVPDSTYDMIGGL  165
Db    86  FVVRIAPWIQRKKLRFGTRVALDQRTMAVVEILPISKDFTVLGFEVIERPNVTVNDIGGL 145

Qy   166  DQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRV 225
```

RESULT 12

```
Db     146  KKQLQELREAIELPLKHPELFEEVGIDPFKGVLLYGPPGCGKTLMAKALAHEVNATFIRV 205

ID   PSMR_PYRAB     STANDARD;      PRT;   399 AA.
AC   Q9V287;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Proteasome-activating nucleotidase (Proteasome regulatory subunit).
GN   PAN OR PYRAB1870 OR PAB2233.
OS   Pyrococcus abyssi.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=29292;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GE5 / Orsay;
RX   PubMed=12622808;
RA   Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
RA   Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA   Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.,
RT   "An integrated analysis of the genome of the hyperthermophilic
RT   archaeon Pyrococcus abyssi.";
RL   Mol. Microbiol. 47:1495-1512(2003).
CC   -!- FUNCTION: Required for the ATP- or CTP-dependent degradation of
CC       proteins, but not small peptides, by the 20S proteasome (By
CC       similarity).
CC   -!- SUBUNIT: Homohexamer (Potential).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
```

```
DR   EMBL; AJ248283; CAB49111.1; -.
DR   PIR; H75207; H75207.
DR   HAMAP; MF_00553; -; 1.
DR   InterPro; IPR005937; 26S_p45.
DR   InterPro; IPR005593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR003960; AAA_sub.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   Proteasome; ATP-binding; Complete proteome.
FT   NP_BIND     181    188       ATP (POTENTIAL).
SQ   SEQUENCE   399 AA;  45350 MW;  DDDBD45915315997C CRC64;

Db   206  VSSELVRKYIGEGARLVHELFELAKEKAPTIIFIDEIDAIGAKRMDETTG-GEREVNRTL 264

Db   265  MQLLAEMDGFDPRGNVKVIAATNRPDILDPALLRPGRFDRLIEVPLPDFEGRLEILKVHT 324

Qy   346  RRMNLMRGIDLKKIAEKMNGASGGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKK 405
Db   325  RRMKL-KGVDLRAIAEMTEGASGADLKAIATEAGMFAIRERRTYVTQEDFLKAVDKVLGN 383

Qy   406  ETEKNMSLRKL 416
Db   384  E------RKL 387
```

Query Match   44.3%; Score 937.5; DB 1; Length 399;

RESULT 13

```
ID   PSMR_ARCFU     STANDARD;      PRT;   398 AA.
AC   O28303;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Proteasome-activating nucleotidase (Proteasome regulatory subunit).
GN   PAN OR AF1976.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC   Archaeoglobaceae; Archaeoglobus.
OX   NCBI_TaxID=2234;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE=98049343; PubMed=9389475;
RA   Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA   Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA   Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA   Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA   Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA   Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA   Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA   Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA   Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA   Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA   Venter J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
CC   -!- FUNCTION: Required for the ATP- or CTP-dependent degradation of
CC       proteins, but not small peptides, by the 20S proteasome (By
CC       similarity).
CC   -!- SUBUNIT: Homohexamer (Potential).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AE000967; AAB89280.1; -.
DR   PIR; G69496; G69496.
DR   HAMAP; MF_00553; -; 1.
DR   TIGR; AF1976.;
DR   InterPro; IPR005937; 26S_p45.
DR   InterPro; IPR005937; -; 1.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR003960; AAA_sub.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   Proteasome; ATP-binding; Complete proteome.
FT   NP_BIND    182    189    ATP (POTENTIAL).
SQ   SEQUENCE   398 AA;  44964 MW;  F3293B7D6A646B4 CRC64;
```

Query Match 43.9%; Score 929; DB 1; Length 398;
Best Local Similarity 52.0%; Pred. No. 9.8e-53;
Matches 193; Conservative 71; Mismatches 98; Indels 9; Gaps 3;

```
Qy    46  LRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEVKVNNGKNKVLVKVHPEGK  105
              |:|   |:|| | |  |
Db    26  IRQELQVRMLEADKERLERELSRLRSEMSRLRQPPAFAGTVIEVLDDDRAIVQNYNGPR   85

Qy   106  YVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEVKVPDSTYDMIGGL 165
Db    86  FVVRIAPMIERDKLRPGARVALDQRTMAIIELLPSSKDPTVLGFEVIERPNVTYNDIGGL  145

Qy   166  DQIKEIKEVIELPIKIPELEFSLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTEIRV   225
Db   146  KKQLQEIREAIELPLKIPELEEVGIDPPKGVLLYGPPGCGKTLMAKALAHEVNATFIRV   205

Qy   226  SGSELVQKYIGEGSRMVRELFVMARENAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTM  285
Db   206  VGSELVRKYIGEGARLVHELFELAKEXAPTIIFIDEIDAIGAKRMDETTG-GEREVNRTL  264

Qy   286  LELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPTPNESSRLDIIKIHS  345
Db   265  MQILAEMDGFDPRGNVKVIAATNRPDILDPALLRPGRFDRLIEVPLPDFEGSRLEILKVHT 324

Qy   346  RRMNLMRGIDLKKIAEKMNGASGGAELKAVCTEAGMFALRERRVHVTCQEDFEMAVAKVMKK 405
Db   325  RRMKL-KNVDLRVIAETEGASGADLKAIATEAGMFAIREERTYVTQEDFLKAVDKVLGN   383

Qy   406  ETEKNMSLRKL  416
Db   384  E-------RKL  387
```

RESULT 14

```
ID   PSMR_METTH     STANDARD;      PRT;   410 AA.
AC   O26824;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Proteasome-activating nucleotidase (Proteasome regulatory subunit).
GN   PAN OR MTH728.
OS   Methanobacterium thermoautotrophicum.
OC   Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC   Methanobacteriaceae; Methanothermobacter.
OX   NCBI_TaxID=187420;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Delta H;
RX   MEDLINE=98037514; PubMed=9371463;
RA   Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
```

Query Match 43.9%; Score 929; DB 1; Length 398;
Best Local Similarity 51.3%; Pred. No. 3.4e-52;
Matches 192; Conservative 64; Mismatches 112; Indels 6; Gaps 2;

```
Qy    20  YYKL------RELYRLEDEKKFIESERIRYEREEVRRLRSEVERLRSPLLLVGVVSDILED  74
Qy    34  YYSLNIHEHQLLRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEVKVNNGK   93
Qy    75  GRVVVKSSTGPKFVVNTSQYINEEELKPGARVALNQQTLAIVNVLPTSKDPMVYGFEVEE  134
Qy    94  NKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEK  153
Qy   154  VPDSTYDMIGGLDQIKEIKEVIELPIKIPELEFSLGIAQPKGVLLYGPPGTGKTLLARA  213
Db   135  KPEVSYED[GGGLDVQIEEIREAVELPLLKPELFAEVGIEPPKGVLLYGPPGTGKTLLAA  194

Qy   214  VAHHTDCTEIRVSGSELVQKYIGEGSRMVRELFVMARENAPLPLIKHPELEFESLGIAQPK 273
Db   195  VANQTRATFIRVSGSEFVQKYIGEGARLVREVFQLAKEXAPSIIFIDELDAIAARRTNSD  254

Qy   274  SGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPTPN  333
Db   255  T-SGDREVQRTMMQLLAELDGFDPRGDVDVIGATNRIDILDPALLRPGRFDRIIEVPLPT  313

Qy   334  EESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGGAELKAVCTEAGMFALRERRVHTQE  393
Db   314  FEGRIQIFKIHTRKMKLAEDVDFKELARITEGASGADIKAICTEAGMFAIRERAKVTML   373

Qy   394  DFEMAVAKVMKKET  407
Db   374  DFTKAIEKVLKKTT  387
```

```
RESULT 15
PSMR_METJA
```

```
Qy   19  NCSAKPTKQGEGLRHYSLNIHEHQLLLRQKTHNLNRLEA----------QRNDLNSRVRMLR    71
            ||  :|  |  || |||  |  :|  | :|::| | ||            :| |:  | |
Db    3  NNSQNVLKKIEDLKKEVRMLKEENSKTKRLMWKIRKLEKDLLIENRKTRLDREVKSLR       62

Qy   72  EELQJLQEPGSYVGEVVKWMGKNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDS    131
Db   63  GEIERRTPLVIATVTEVLDDHRVAVKSTTGPHRVINYSRFIDRKQLEPGARVALNQQT      122

Qy  132  YVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQOIKEIKEVIELPIKHPELFSSLGI    191
            |:|:|||: :||:| :| ||:|:| : |:|||| :: ::|:||:|||:| |||| :|||
Db  123  FSIVDVLPSEKDPVVTGMEVEEKDVSYEQIGGLEEQVREVKTVELPLKKPELFEKIGI      182

Qy  192  AQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKIIGEGSRMVRELFVMARE    251
            : |||||||||||||||||:|||||: : :||::|||:|:::||||| ::|||:|: |:|
Db  183  EPPKGVLLYGPPGTGKTLLAKAVAHETNAFFIKIVASEFVRKYIGEGARLVRGVFELAKE    242

Qy  252  HAPSIIFMDEIDSIGSARMESSGGSGNGDSEVQRTMLELINQLDGFEASNKIKVLMATNRID   311
            ::|||||:|||| : :|:::| | : ::| ||||::: |::::|||||: :  :::|||||:|
Db  243  KSPSIIFIDEIDAVAAKRLKS-SISGDREVQRTLMQLLAELDGFESRGNVGIVAATNRPD     301

Qy  312  ILDQALLRPGRIDRKIEFPPPNEESRLDILKIHSRMNLMRGIDLKKIAEKMNGASGAEL    371
            |||:||||||| || ||:| ||:: | |:||| | |  ::  :| ::| | ||||||:|
Db  302  ILDPALLRPGRFDRFIEVPLPNEDGRRELKIHTSGMALAEEVDIELLARITDGASGADL    361

Qy  372  KAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEK   409
            ||:|||||||| | :  :||:|| :|| ||:|  |:|:
Db  362  KAICTEAGMFAIRDEDEVTMADFMDAVDKIMGVEKEE    399
```

```
Query Match      43.6%;  Score 923;  DB 1;  Length 410;
Best Local Similarity   47.5%;
Matches 189;  Conservative   80;  Mismatches 121;  Indels    8;  Gaps    2;
```

Left record:

```
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
CC   -!- FUNCTION: Required for the ATP- or CTP-dependent degradation of
CC       proteins, but not small peptides, by the 20S proteasome (by
CC       similarity).
CC   -!- SUBUNIT: Homohexamer (Potential).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------
DR   EMBL; AE000852; AAB85233.1; -.
DR   PIR; C69197; C69197.
DR   HAMAP; MF_00553; -; 1.
DR   InterPro; IPR005937; 26S_p45.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR003960; AAA_sub.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   Proteasome; ATP-binding; Complete proteome.
FT   NP_BIND      192    199    ATP (POTENTIAL).
SQ   SEQUENCE   410 AA;  46111 MW;  38261703F836EBAD4 CRC64;
```

Right record:

```
ID   PSMR_METJA     STANDARD;      PRT;    430 AA.
AC   Q58576;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Proteasome-activating nucleotidase (Proteasome regulatory subunit).
GN   PAN OR MJ1176.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96379999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=20158885; PubMed=10922374;
RA   Wilson H.L., Ou M.S., Aldrich H.C., Maupin-Furlow J.;
RT   "Biochemical and physical properties of the Methanococcus jannaschii
RT   20S proteasome and PAN, a homolog of the ATPase (Rpt) subunits of the
RT   eucaryal 26S proteasome.";
RL   J. Bacteriol. 182:1680-1692(2000).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=99403035; PubMed=10473546;
RA   Zwickl P., Ng D., Woo K.M., Klenk H.-P., Goldberg A.L.;
RT   "An archaebacterial ATPase, homologous to ATPases in the eukaryotic 26
RT   S proteasome, activates protein breakdown by 20 S proteasomes.";
RL   J. Biol. Chem. 274:26008-26014(1999).
CC   -!- FUNCTION: Required for the ATP- or CTP-dependent degradation of
CC       proteins, but not small peptides, by the 20S proteasome.
CC   -!- COFACTOR: MAGNESIUM.
CC   -!- ENZYME REGULATION: INHIBITED BY EDTA, NEM AND PCMS.
CC   -!- SUBUNIT: Homohexamer (Potential).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- MISCELLANEOUS: ITS OPTIMUM PH IS 7-8 AND THE OPTIMUM TEMPERATURE
CC       IS 80 DEGREES CELSIUS.
CC   -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------
DR   EMBL; U67559; AAB99179.1; -.
DR   PIR; G64446; G64446.
DR   HAMAP; MF_00553; -; 1.
DR   TIGR; MJ1176; -.
DR   InterPro; IPR005937; 26S_p45.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR003960; AAA_sub.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   Proteasome; ATP-binding; Complete proteome.
FT   NP_BIND      211    218    ATP (POTENTIAL).
```

```
SQ   SEQUENCE   430 AA;  48689 MW;  3FD2E94A68D4B3DD CRC64;

     Query Match         42.8%;    Score 906;  DB 1;  Length 430;
     Best Local Similarity   45.0%;    Pred. No. 1.1e-50;
     Matches 186;  Conservative 86;  Mismatches 133;  Indels   8;  Gaps   2;

Qy       4  VGVELKHAAEGVPEANCSAKPTKQGEGLRHYSLNIHEHQLLLRQKTHNLNRLEAQRNDL  63
            :  |||  :        :      :                  |||   ||: ||
Db       7  ISTELKKEKAAFTEEFKEKEINDNSNLKNDLLKEELQEKARIAELESRILKLBLEKKEL  66

Qy      64  N--------SRVRMLREELQLLQEPGSYVGEVKVKVMGKNKVLVKVHPEGRKYVVDIDKNIDI  116
                     :  |: ||||:  |  || |||:|   ::|
Db      67  ERENLQLMKENEILRRELDRMRVPPLIVGTVVDKVGERKVVVKSSTGPSFLVNVSHFVNP  126

Qy     117  TKITPSTRVALRNDSVVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVI  176
                | |  :                  ||| ::|   :|: :: ::|:|||:
Db     127  DDLAPGKRVCLNQQTLTVVDVLPENKDYRAKAMEVDERPNVRYEDIGGLEKQMQEIREVV  186

Qy     177  ELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKYIG  236
            |||:||||||:  | : |   ||||||||||||||||  |  :| |||:| ||:|:|| :
Db     187  ELPLKHPELFEKVGIEPPKGILLYGPPGTGKTLLAKAVATETNATFIRVVGSELVKKFIG  246

Qy     237  EGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFE  296
            ||:|||||||  :   |||||||||||| :|| |  :  | |:|||||:|||||:|||| :
Db     247  EGASLVKDIFKLAKEKAPSIIFIDEIDAIAAKRTDALTG-GDREVQRTLMQLLAEMDGFD  305

Qy     297  ASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPTPNEESRLDILKIHSRRMNLMRGIDL  356
            : :: ||| || || | || ::||  | ||:  |   :| ||:: |:| | ||::     :
Db     306  ARGDVKIIGATNRPDILDPAILRPGRFDRIIEVPAPDEKGRLEILKIHTRKMNLAEDVNL  365

Qy     357  KKIAEKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEK  409
            ::||  |  :  ||||:|  ||||:||:  |:|| |||||:| | :|:|::|
Db     366  EEIAKMTEGCVGAELKAICTEAGMNAIRELRDYVTMDDFRKAVEKIMEKKKVK  418
```

Search completed: November   9, 2003, 12:46:09
Job time : 25 secs

OM protein - protein search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: November 9, 2003, 12:29:56 ; Search time 110 Seconds
                 (without alignments)
                 980.600 Million cell updates/sec

Title: US-09-462-972-2
Perfect score: 2116
Sequence: 1 MALVGVELKHAAEGVPEANC..........VAKVMKKETEKNMSLRKLWK 418

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 listing first 45 summaries

Database :   SPTREMBL_23:*
                 1: sp_archea:*
                 2: sp_bacteria:*
                 3: sp_fungi:*
                 4: sp_human:*
                 5: sp_invertebrate:*
                 6: sp_mammal:*
                 7: sp_mhc:*
                 8: sp_organelle:*
                 9: sp_phage:*
                 10: sp_plant:*
                 11: sp_rodent:*
                 12: sp_virus:*
                 13: sp_vertebrate:*
                 14: sp_unclassified:*
                 15: sp_rvirus:*
                 16: sp_bacteriap:*
                 17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1926.5	91.0	419	10	Q9C5U3	q9csu3 arabidopsis
2	1913	90.4	405	10	Q9SEI1	q9sei1 arabidopsis
3	1905.5	90.1	419	10	Q9SEI1	q9sei1 arabidopsis
4	1901.5	89.9	424	10	Q94BQ2	q94bq2 arabidopsis
5	1896	89.6	424	10	Q9M613	q9m613 pinus taeda
6	1879.5	88.8	414	10	Q9FEB6	q9feb6 oryza sativ
7	1872.5	88.5	414	10	Q9SST5	q9sst5 matricaria
8	1861	87.9	452	10	Q9SCD6	q9scd6 fagus sylva
9	1673	79.1	406	11	Q9FNU8	q9fnu8 dactylis gl
10	1622.5	76.7	399	5	Q9CWN5	q9cwn5 mus musculu
11	1588.5	75.1	435	5	Q9VA54	q9va54 drosophila
12	1572	74.3	416	5	Q8IAU5	q8iau5 plasmodium
13	1528.5	72.2	389	5	Q9XTT9	q9xtt9 caenorhabdi
14	1501	70.9	411	5	Q9P665	q9p665 neurospora
15	1501	70.9	432	5	Q8MXF1	q8mxf1 caenorhabdi
16	1420	67.1	453	5	Q8SQK0	q8sqk0 encephalito
17	1378	65.1	408	5	Q9NC95	q9nc95 trypanosoma
18	1365.5	64.5	306	11	Q99KR1	q99kr1 mus musculu
19	1327	62.7	348	11	Q8K1K2	q8k1k2 mus musculu
20	1257.5	59.4	280	4	O43208	o43208 homo sapien
21	1076	50.9	400	8	Q98S38	q98s38 guillardia
22	932.5	44.1	217	10	Q9GGW9	q9ggw9 amblyomma a
23	895	42.3	439	10	Q9VCG1	q9vcg1 drosophila
24	888	42.0	475	11	Q8BVQ9	q8bvq9 mus musculu
25	884	41.8	433	5	Q9V478	q9v478 drosophila
26	877.5	41.5	390	5	Q8SRM0	q8srm0 encephalito
27	869.5	41.1	455	5	Q8IJW0	q8ijw0 plasmodium
28	868.5	41.0	390	5	Q9W414	q9w414 drosophila
29	867.5	41.0	399	10	Q9SEI3	q9sei3 arabidopsis
30	867	41.0	390	10	Q9U7A2	q9u7a2 drosophila
31	867	41.0	389	11	Q9CXH9	q9cxh9 mus musculu
32	865.5	40.9	451	10	Q9SSB4	q9ssb4 arabidopsis
33	862	40.7	443	10	Q9FT10	q9ft10 brassica na
34	861.5	40.7	397	5	Q8SZI9	q8szi9 drosophila
35	860	40.6	393	5	Q8IEQ1	q8ieq1 plasmodium
36	860	40.6	399	10	Q9MAK9	q9mak9 arabidopsis
37	859.5	40.6	443	10	Q9SL67	q9sl67 arabidopsis
38	858.5	40.6	398	10	Q9SEI5	q9sei5 arabidopsis
39	858	40.6	443	10	Q9SZD4	q9szd4 arabidopsis
40	857.5	40.6	396	5	062556	062556 manduca sex
41	857	40.5	398	5	Q9VTQ9	q9vtq9 drosophila
42	856.5	40.5	415	5	Q9SS82	q9ss82 arabidopsis
43	856.5	40.5	459	3	Q9P723	q9p723 neurospora
44	855.5	40.4	398	10	Q8H9D4	q8h9d4 solanum tub
45	855.5	40.4	443	10	Q94JT0	q94jt0 arabidopsis

ALIGNMENTS

RESULT 1
Q9C5U3

ID   Q9C5U3   PRELIMINARY;   PRT;   419 AA.
AC   Q9C5U3;
DT   01-JUN-2001 (TREMBLrel. 17, Created)
DT   01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT   01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE   AtSUG1 protein (Hypothetical 47.2 kDa protein).
GN   AtSUG1.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA   Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA   Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA   Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA   Ecker J., Theologis A., Davis R.W.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kuromori T., Yamamoto M.;
RT   "Identification of a cDNA from Arabidopsis thaliana for a member of
RT   conserved Sug1 CAD family.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB044348; BAB40755.1; -.
DR   EMBL; AY065174; AAL38350.1; -.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR005937; 26S_P45.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR003960; AAA_ATPase_sub.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26SP45; 1.
DR   PROSITE; PS00674; AAA; 1.

```
KW   Hypothetical protein; ATP-binding.
SQ   SEQUENCE   419 AA;  47248 MW;  3186E05B7142EC8B CRC64;

Query Match      91.0%;      Score 1926.5;  DB 10;  Length 419;
Best Local Similarity    92.1%;   Pred. No. 6.6e-117;
Matches  387;  Conservative  10;  Mismatches  20;  Indels  3;  Gaps  2;

Qy        1 MALVGVELKHAAEGVPEANCSAK--PTKQGEGLRHYSLNIHEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLREEL 74
            |||||:|.|      |:.: ||||  ||||||||||||||||||||||:.||||||||||||||||||||||||||
Db        2 EETCNVKGAAAKQGEGLKQYLQHIHELQRQLRQKTKTNNLNRLEAQRNELNSRVRMLREEL 61

Qy        1 MAAVGVDSRR-PETAMEETCNVKGAAAKQGEGLKQYLQHIHELQRQLRQKTKTNNLNRLEA 59
            |:|||||.|| ||:.||||||||||||||||||||||||||||||||||||||||||||||
Db       62 QLLQEPGSYGEVVKVKVMGKNKVLVKVKVHPEGKYVVVDIDKNIDITKITPSTRVALRNDSYVL 121

Qy        1 QRNDLNSRVRMLREELQLLQEPGSYGEVVKVKVMGKNKVLVKVKVHPEGKYVVVDIDKNIDITK 58
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db       75 QLLQEPGSYGEVVKVKVMGKNKVLVKVKVHPEGKYVVVDIDKNIDITKITPSTRVALRNDSYVL 134

Qy       59 QRNELNSRVRMLREELQLLQEPGSYGEVAVKVMGKNKVLVKVHPEGKYVVVDIDKSIDITK 119
            |||||||||||||||||||||||||||:|||||||||||||:||||||||||||:||||||
Db      122 HLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQIKEIKEVIEL 181

Qy      119 ITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIEL 178
            ||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db      182 KGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKYIGEG 241

Qy      120 ITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIEL 179
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      195 KGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKYIGEG 254

Qy      179 PIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKYIGEG 238
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      255 SIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILD 314

Qy      180 PIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKYIGEG 239
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      182 KGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKYIGEG 241

Qy      239 SRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEAS 298
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      242 SIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILD 301

Qy      240 SRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEAS 299
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      315 QALLRPGRIDRKIEFPNEESRFDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAV 374

Qy      299 NKIKVLMATNRIDILDQALLRPGRIDRKIEFPTNEESRLDILKIHSRRMNLMRGIDLKK 358
            |||||||||||||||||||||||||||||||::||||:|||||||||||||||||||||
Db      302 QALLRPGRIDRKIEFPNEESRFDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAV 361

Qy      300 NKIKVLMATNRIDILDQALLRPGRIDRKIEFPNPNEESRFDILKIHSRRKNMLRGIDLKK 359
            ||||||||||||||||||||||||||||||||::||||:|||||||||||:|||||||||
Db      362 CTEAGMFALRERRVHVTQEDFEMAVAKVMKKDTEKNMSLRKLWK 405

Qy      359 IAEKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
            |||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db      360 IAEKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKDTEKNMSLRKLWK 419

Qy      359 IAEKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
            |||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db      360 IAEKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKDTEKNMSLRKLWK 419
```

```
RESULT 2
ID   Q9SEI1          PRELIMINARY;      PRT;    405 AA.
AC   Q9SEI1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   26S proteasome AAA-ATPase subunit RPT6a.
GN   RPT6A.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Fu. H., Doelling J.H., Rubin D.M., Vierstra R.D.;
RT   "Structure and function analysis of the six regulatory particle
RT   triple-A ATPase subunits RPT1-6 of the 26S proteasome from Arabidopsis
RT   thaliana.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF123395; AAF22526.1; -.
DR   InterPro; IPR005937; 26S_P45.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase.
DR   InterPro; IPR003960; AAA_ATPase_centr.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   ATP-binding; Proteasome.
SQ   SEQUENCE   405 AA;  45806 MW;  1349782122A0171D CRC64;

Query Match      90.4%;      Score 1913;  DB 10;  Length 405;
Best Local Similarity    94.3%;   Pred. No. 4.7e-116;
Matches  381;  Conservative  8;  Mismatches  13;  Indels  2;  Gaps  1;
```

```
Qy       17 EANCSAK--PTKQGEGLRHYSLNIHEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLREEL 74
            |||||||  |||||:||.|||||||||||||.||:|||||||||||:.||||||||||||
Db        2 EETCNVKGAAAKQGEGLKQYLQHIHELQRQLRQKTKTNNLNRLEAQRNELNSRVRMLREEL 61

Qy       62 QLLQEPGSYGEVVKVKVMGKNKVLVKVKVHPEGKYVVVDIDKNIDITKITPSTRVALRNDSYVL 121
            ||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||
Db       75 QLLQEPGSYGEVVKVKVMGKNKVLVKVKVHPEGKYVVVDIDKNIDITKITPSTRVALRNDSYVL 134

Qy      122 HLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQIKEIKEVIELPIKHPELFESLGIAQP 181
            |||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db      135 HLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQIKEIKEVIELPIKHPELFESLGIAQP 194

Qy      195 KGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKYIGEGSRMVRELFVMAREHAP 254
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      182 KGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKYIGEGSRMVRELFVMAREHAP 241

Qy      182 KGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKYIGEGSRMVRELFVMAREHAP 241
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      255 SIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILD 314

Qy      255 SIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILD 314
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      242 SIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILD 301

Qy      315 QALLRPGRIDRKIEFPTNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAV 374
            ||||||||||||||||:||||:|||||||||||||||||||||||||||||||||||||
Db      302 QALLRPGRIDRKIEFPNEESRFDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAV 361

Qy      302 QALLRPGRIDRKIEFPNEESRFDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAV 361
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      375 CTEAGMFALRERRVHVTQEDFEMAVAKVMKKDTEKNMSLRKLWK 418

Qy      375 CTEAGMFALRERRVHVTQEDFEMAVAKVMKKDTEKNMSLRKLWK 418
            ||||||||||||||||||||||||||||||||||||||||||||
Db      362 CTEAGMFALRERRVHVTQEDFEMAVAKVMKKDTEKNMSLRKLWK 405
```

```
RESULT 3
ID   Q94BQ2          PRELIMINARY;      PRT;    419 AA.
AC   Q94BQ2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Putative 26S proteasome AAA-ATPase subunit RPT6a.
GN   At5G20000.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA   Dale J.M., Gibson H.A., Goldsmith A.D., Jiang P.X., Lee J.M.,
RA   Quach H.L., Tang C., Toriumi M., Yu G., Bowser L., Carninci P.,
RA   Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
RA   Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B., Lin J.,
RA   Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
RA   Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT   "Full length cDNA of gene F28I16_150/At5g20000.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Etgu P., Lee J.M.,
RA   Toriumi M., Yu G., Brooks S., Chao Q., Chen H., Karlin-Neumann G.,
RA   Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,
RA   Southwick A., Davis R.W., Ecker J.R., Theologis A.;
RT   "Arabidopsis Open Reading Frame (ORF) Clones.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA   Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT   "Full-length messenger RNA sequences greatly improve genome
RT   annotation.";
RL   Genome Biol. 0:0-0(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
```

```
RA   Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA   Feldmann K.;
RT   "Full-Length cDNA from Arabidopsis thaliana.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY039965; AAK64142.1; -.
DR   EMBL; AY039403; AAL85134.1; -.
DR   EMBL; AY087503; AAM65046.1; -.
DR   InterPro; IPR005937; 26S_P45.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR003960; AAA_sub.
DR   Pfam; PF00004; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   Proteasome.
SQ   SEQUENCE   419 AA;  47156 MW;  B464F8811048F1F CRC64;

Query Match             90.1%;  Score 1905.5;  DB 10;  Length 419;
Best Local Similarity   90.9%;
  Matches  381;  Conservative   12;  Mismatches   25;  Indels    1;  Gaps    1;

Qy    1 MALVGVELKHAAEGVPE-ANCSAKPTKQGEGLRHYYSLNIHEHOLLLRQKTHNINRLEAQ 59
Db    1 MAAVGVEARPPVTAMEETCNVKGAAAKQGEGLNKYYLQHLDELQRLQREBKSYNLNRLEAQ 60

Qy   60 RNDLNSRVRMLREEOLQLLQEPGSYGVEVVKVKMGKQKNVLVLVKVHPEGKYVVDIDKNIDITKI 119
Db   61 RNELNSRVRMLREELQLLQEPGSYGVEVVKVKWGKNKVLVKVHPEGKYVVDIKSIDITKL 120

Qy  120 TPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELP 179
Db  121 TPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELP 180

Qy  180 IKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKYIGEGS 239
Db  181 IKHPELFESLGTAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSEL 240

Qy  240 RVMRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASN 299
Db  241 RVMRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASN 300

Qy  300 KIKVLMATNRIDILDQALLRPGRIDRKIEFPTPNEESRLDILKIHSRRMNLMRGIDLKKI 359
Db  301 KIKVLMATNRIDILDQALLRPGRIDRKIEFPNPNEESRPDILKIHSRKWNLMRGIDLKKI 360

Qy  360 AEKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
Db  361 AEKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVWKKDTEKNMSLRKLWK 419
```

```
RESULT 4
Q9M613
ID   Q9M613
AC   Q9M613;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   26S proteasome regulatory subunit 8.
OS   Pinus taeda (Loblolly pine).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX   NCBI_TaxID=3352;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=260; TISSUE=Somatic embryo;
RA   Destefano-Beltran L.J.C., Casas-Mollano A., Cairney J.;
RT   "Isolation of a full-length cDNA encoding a putative 26S proteasome
RT   subunit 8 from Pinus taeda somatic embryos.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
     EMBL; AF220199; AAF27916.1; -.
DR   InterPro; IPR005937; 26S_P45.
DR   InterPro; IPR005593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR003960; AAA_sub.

PRELIMINARY;     PRT;     433 AA.
```

```
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   ATP-binding; Proteasome.
SQ   SEQUENCE   433 AA;  48931 MW;  05C32ABB05047070 CRC64;

Query Match             89.9%;  Score 1901.5;  DB 10;  Length 433;
Best Local Similarity   89.0%;
  Matches  381;  Conservative   14;  Mismatches   20;  Indels   13;  Gaps    1;

Qy    4 VGVELKHAAEGVPEANC-----------SAKPTKQGEGLRHYYSLNIHEHOLLLRQKT 50
Db    6 LGGILKPASDLEAAANAKEKLKEKHQLPKESNRMVKHGEGLRQYYLQHIHDLQLQVRQKT 65

Qy   51 HNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYGVEVVKVKMGKNKVLVKVHPEGKYVVDI 110
Db   66 HNLNRLEAQRNDLNSKVRALREELQLLQEPGSYGVEVVKVMGKSKVLVKVHPEGKYVVDV 125

Qy  111 DKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIK 170
Db  126 DKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIK 185

Qy  171 EIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSEL 230
Db  186 EIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSEL 245

Qy  231 VQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLN 290
Db  246 VQKIIGEGSSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGTGNGDSEVQRTMLELLN 305

Qy  291 QLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPTPNEESRLDILKIHSRRMNL 350
Db  306 QLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPNPNEESRFDILKIHSRKMNL 365

Qy  351 MRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKN 410
Db  366 MRGTDLKKIAEKMSGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVWKKETEKN 425

Qy  411 MSLRKLWK 418
Db  426 MSLRKLWK 433
```

```
RESULT 5
Q9FEB6
ID   Q9FEB6
AC   Q9FEB6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   26S proteasome ATPase subunit Rpt6 (26S proteasome regulatory particle
DE   triple-A ATPase subunit6).
GN   ORSPT6.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yamamoto T., Kimura S., Oka M., Ishibashi T., Yanagawa Y., Nara T.,
RA   Nakagawa H., Hashimoto J., Sakaguchi K.;
RT   "26S proteasome ATPase subunit.";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Shibahara T., Kawasaki H., Hirano H.;
RT   "Structural analysis of the regulatory particle triple-A ATPase
RT   subunits from the Rice 26S proteasome.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB052887; BAB19880.1; -.

PRELIMINARY;     PRT;     424 AA.
```

```
DR   EMBL; AB033537; BAB17626.1; -.
DR   Gramene; Q9FEB6; -.
AC   Q9FEB6;
DR   InterPro; IPR005937; 26S_P45.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR003960; AAA_sub.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   ATP-binding; Proteasome.
SQ   SEQUENCE   424 AA;   47222 MW;  49197D5C6A584180 CRC64;

Query Match       89.6%;  Score 1896;  DB 10;  Length 424;
Best Local Similarity  95.4%;  Pred. No. 6.3e-115;
Matches  373;  Conservative  11;  Mismatches  7;  Indels  0;  Gaps  0;

Qy       28  GEGLRHYYSLNIHIEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEV  87
             |||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       34  GEGLRQYYLQHIHDLQLQIRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEV  93

Qy       88  VKVMGKNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVN  147
             |||||:|||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db       94  VKVMGKSKVLVKVHPEGKYVVDIDKSIDITKITPSTRVALRNDSYMLHLILPSKVDPLVN  153

Qy      148  LMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGK  207
             |||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      154  LMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGK  213

Qy      208  TLLARAVAHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGS  267
             |:||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db      214  TLLARAVAHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGS  273

Qy      268  ARMESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKI  327
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      274  ARMESGTGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKI  333

Qy      328  EFPTPNEEESRLDILKIIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERR  387
             ||:|||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      334  EFPNPNEDSRFDILKTIHSRKMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERR  393

Qy      388  VHVTQEDFEMAVAKVMKKETEKNMSIRKLWK  418
             |||||||||||||||||||||||||||||||
Db      394  VHVTQEDFEMAVAKVMKKDTEKNMSLRKLWK  424
```

```
RESULT 6
Q9SST5
ID   Q9SST5       PRELIMINARY;    PRT;   414 AA.
AC   Q9SST5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   TAT-binding protein homolog.
GN   MCTBP1.
OS   Matricaria chamomilla.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
OC   Anthemideae; Matricaria.
OX   NCBI_TaxID=98504;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ashida Y., Nishimoto M., Hirata T.;
RT   "TAT-binding protein homolog, Matricaria chamomilla.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB035272; BAA87070.2; -.
DR   InterPro; IPR005937; 26S_P45.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR003960; AAA_sub.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
```

```
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   ATP-binding.
SQ   SEQUENCE   414 AA;   46269 MW;  AE2F49C09E8B6E52 CRC64;

Query Match       88.8%;  Score 1879.5;  DB 10;  Length 414;
Best Local Similarity  93.0%;  Pred. No. 7.2e-114;
Matches  373;  Conservative  13;  Mismatches  14;  Indels  1;  Gaps  1;

Qy       19  NCSAK-PTKQGEGLRHYYSLNIHIEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLREELQLL  77
             |:|:| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       14  SCSSKGVAKTGEGLRQYYLQHIHDAQLQVRQKTHNLNRLEAQRNELNSRVGMLKEELQLL  73

Qy       78  QEPGSYVGEVVKVMGKNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLV  137
             ||||||||||||||||:|||||||||||||||||||:||||||||||||||||||||||||
Db       74  QEPGSYVGEVKVMGKSKVLVKVHPEGKYVVDIDKSIDITKITPSTRVALRNDSYVLHLV  133

Qy      138  LPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGV  197
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      134  LPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGV  193

Qy      198  LLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSII  257
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      194  LLYGPPGTGKTLLARAVAHHTDCAFIRVSGSELVQKYIGEGSRMVRELFVMAREHAPSII  253

Qy      258  FMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQAL  317
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      254  FMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDPAL  313

Qy      318  LRPGRIDRKIEFPTPNEEESRLDILKIIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTE  377
             ||||||||:|||||||||:||||||||||||||||||||||||||||||||||||||||||
Db      314  LRPGRIDREIEFPNPNEDSRFDILKTIHSRKMNLMRGIDLKKIAEKMNGASGAELKAVCTE  373

Qy      378  AGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSIRKLWK  418
             |||||||||||||||||||||||||||||:||||||||||
Db      374  AGMFALRERRVHVTQEDFEMAVAKVMKKDTEKNMSLRKLWK  414
```

```
RESULT 7
Q9SCD6
ID   Q9SCD6       PRELIMINARY;    PRT;   418 AA.
AC   Q9SCD6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   26S proteasome subunit 8.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Seed;
RA   Lorenzo O., Rodriguez D., Nicolas G., Nicolas C.;
RT   "Up-regulation by GA3 of a new member of the AAA family (FsA1) in
RT   dormant beechnuts (Fagus sylvatica L.) showing Mg2+-dependent ATPase
RT   activity.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ251819; CAB63651.1; -.
DR   InterPro; IPR005937; 26S_P45.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   ATP-binding; Proteasome.
SQ   SEQUENCE   418 AA;   47423 MW;  B6CBB136F4FD51CA CRC64;

Query Match       88.5%;  Score 1872.5;  DB 10;  Length 418;
Best Local Similarity  90.7%;  Pred. No. 2.1e-113;
Matches  380;  Conservative  11;  Mismatches  23;  Indels  5;  Gaps  5;
```

```
OS   Fagus sylvatica (Beechnut).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fagales; Fagaceae; Fagus.
OX   NCBI_TaxID=28930;
GN   Al.
DE   26S proteasome subunit 8.
```

```
Qy      4  VGVELKHAAEGVP-EANCSAKPTKQGEGLRHYYSLNIHEHQLLRQKTHNLRLEAQRND  62
              :|||:|||||| |||||||||||||||| :||||||||||||||||||||||:|||
Db      1  MAIETKH-PEAVPADETCSAKCSKQGEGLRHYYSLNIHEHQLQVRQKTHNLRLEAQRNE  59

Qy     63  LNSRVRMLREELQLLQEPGSYVGEVVKVKVMGKNKVLVKVHPEGKYVVVDIDKNIDITKITPS  122
              |||||||||||||||||||||||||||||||||||||||||||||:||:|||||||||
Db     60  FNSRVRMLREELQLLQEPGSYVGEVVKVKVMGKNKVLVKVHSEGKYVVVDIDKNIDITKITPS  119

Qy    123  TRVALRNDSYVLHLVLPSKVDPLVNLMKYEKVPDSTYDMIGGLDQQIKEIKEVIELPIKH  182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db    120  TRVALRNDSVLHLLLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKH  179

Qy    183  PELF-ESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKIIGEGSRM  241
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    180  PELFLESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKIIGEGSRM  239

Qy    242  VRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNKI  301
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    240  VRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQDHMLELLNQLDGFESIKQI  299

Qy    302  KVLMATNRIDILDQALLRPGRIDRKIEFPTPNEESRLDILKIHSRRMNLMRGIDLKKIAE  361
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    300  KVLMRTNRIDILDQALLRPGRIDRKIEFPNPNRESRFDILKIHSRRMNLMRGIDLKKIAE  359

Qy    362  KMNGASGAELKAVCTEAGMPFAL-RERRVH-VTQEDFEMAVAKVMKKETEKNMSLRKLWK  418
              ||||||||||||||||||||||| |||||| |||||||||||||:|||||||||||||
Db    360  KMNGASGAELKAVCTEAGMFALTAERRVRLVTQEDFEMAVAKVMKKETEKNMSLRKLWK  418
```

RESULT 8

ID   Q9FNU8   PRELIMINARY;   PRT;   452 AA.
AC   Q9FNU8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   26S proteasome RPT6a subunit.
GN   DGRPT6A.
OS   Dactylis glomerata (Orchard grass) (Cocksfoot grass).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC   Poeae; Dactylis.
OX   NCBI_TaxID=4509;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryogenic leaf;
RA   Alexandrova K.S., Conger B.V.;
RT   "Identification of a cDNA clone for orchardgrass homolog of the 26S
RT   proteasome RPT6a subunit.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY011124; AAG42150.1; -.
DR   InterPro; IPR005937; 26S_p45.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR003960; AAA_sub.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   ATP-binding; Proteasome.
SQ   SEQUENCE   452 AA; 50367 MW; 4BFEA64294D99938 CRC64;

Query Match   87.9%; Score 1861; DB 10; Length 452;
Best Local Similarity   89.1%; Pred. No. 1.3e-112;
         Conservative 16; Mismatches 22; Indels 8; Gaps 3;

Matches 376;

```
Qy      1  MALVGVELKH-----AAEGVPEANCSAKPTKQ-GEGLRHYYSLNIHEHQLLRQKTHNLRL  56
              |||| ||::|||     :||:|||||||||||| ||||||||||||||:||:|||||||||
Db     35  MATVAMDISKPPPAAG----DDSAKGARAGGEGLRQYYLQHIHDLQIRTKIHNLNRL  90

Qy     57  EAQRNDLNSRVRMLREELQLLQEPGSYVGEVVKVKVMGKNKVLVKVHPEGKYVVVDIDKNIDI  116
              ||||::|:|:|||||||||||||||||||||||||||||||||||||||:|||||||||:|||
```

RESULT 9

ID   Q9CWN5   PRELIMINARY;   PRT;   406 AA.
AC   Q9CWN5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Protease (Prosome, macropain) 26S subunit, ATPase 5.
GN   PSMC5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryonic stem cells;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK010505; BAB26990.1; -.
DR   MGD; MGI:1056047; Psmc5.
DR   InterPro; IPR005937; 26S_p45.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR003960; AAA_sub.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.

```
Db     91  EAQRNDLNSKVRALREELQLLQEPGSYVGEVVKVKVMGKSKVLVKVHPEGKYVVDIDKSIDI  150

Qy    117  TKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMKYEKVPDSTYDMIGGLDQQIKEIKEVI  176
              ||||||||||||||:||:|:||||||||||||||||||||||||||||||||||||||||
Db    151  TKITPSTRVALRNDSYMLHLLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVI  210

Qy    177  ELPIKHPELFLESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKYIG  236
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db    211  ELPIKHPELFLESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKYIG  270

Qy    237  EGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFE  296
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    271  EGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFE  330

Qy    297  ASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPTPNEESRLDILKIHSRRMNLMRGIDL  356
              |||||||||||||||||||||||||||||||||||:|||:|||||||||||||||||||
Db    331  ASNKIKVLMATNRIDILDQALLRPGRIDRKIEFPNPNEDSRGDILKIHSTTMNLMRGIDL  390

Qy    357  KKIAEKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKL  416
              ||||||||||||||||||||||||||:||||||||||||||||||| ||||||||||||
Db    391  KKIAGKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKDNEKNMSLRKL  450

Qy    417  WK 418
              || ||
Db    451  WK 452
```

```
DR   PROSITE; PS00674; AAA; 1.
KW   ATP-binding;
SO   SEQUENCE   406 AA;  45527 MW;  5F1D310C4A85A7E3 CRC64;
     Query Match             Best Local Similarity  79.1%;
     Conservative 37;  Matches 330;  Mismatches 39;
     Score 1673;  DB 11;  Length 406;  Pred. No. 1.7e-100;
     Indels 12;  Gaps 2;

Qy    1  MALVGVELKHAAEGVPEANCSAKPTXQGEGLRHYYSLNIHEHQLLIRQKTHNLNRLEAQR  60
Db    1  MALDGPEQMELEEG--------------------KAGSGLRQYLSKIEELQLIVNDKQQNLRLQAQR  49

Qy   61  NDLNSRVRMLREELQLLQEPGSSYVGEVAVKVNGKNKVLVKHPEGKYVVDIDXNIDITKIT  120
Db   50  NELNAKVRLLREELQLLQEQGSSYVGEVAVRAMDKKKVLVKVHPEGKRVVVDDVDKNIDNDVT  109

Qy  121  PSTRVALRNDSYVLHJVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPI  180
Db  110  PNCRVALRNDSYTLKILPNKVDPLVSLMMVEKVPDSTYEMIGGILDKQIKEIKEVIELPV  169

Qy  181  KHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKYIGEGSR  240
Db  170  KHPELFEALGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKFIGEGAR  229

Qy  241  MVRELFVMAREHAPSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNK  300
Db  230  MVRELFVMAREHAPSIIFMDEIDSIGSSRLEGGSG-GDSEVQRTMLELLNQLDGFEATKN  288

Qy  301  IKVLMATNRIDILDLQALLRPGRIDRKIEFPPPNEESRLDILKIHSRRMNLMRGIDLKKIA  360
Db  289  IKVIMATNRIDILDSALLRPGRIDRKIEFPPPNEEARLDILKIHSRKMNLTRGINLRKIA  348

Qy  361  EKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK  418
Db  349  ELMPGASGAEVKGVCTEAGMYALRERRVHVTQEDFEMAVAKVMQKDSEKNMSIKKLWK  406


RESULT 10
SEQUENCE FROM N.A.

ID   Q9VA54      PRELIMINARY;      PRT;    399 AA.
AC   Q9VA54;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CG2241 protein (SD17676p).
GN   CG2241
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003774; AAF57069.1; -.
DR   EMBL; AY119222; AAM51089.1; -.
DR   FlyBase; FBgn0039788; CG2241.
DR   InterPro; IPR005937; 26S_p45.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR003960; AAA_sub.
DR   Pfam; PF00004; AAA; 1.
DR   PROSITE; PS00674; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.
KW   ATP-binding;
SO   SEQUENCE   399 AA;  45169 MW;  F880C1C2D0EACBF6 CRC64;
     Query Match             Best Local Similarity  77.8%;
     Conservative 45;  Matches 315;  Mismatches 34;
     Score 1622.5;  DB 5;  Length 399;  Pred. No. 3e-97;
     Indels 11;  Gaps 2;

Qy   14  GVPEANCSAKPTXQGEGLRHYYSLNIHEHQLLIRQKTHNLNRLEAQRNDLNSRVRMLREE  73
Db    6  GIPEGS-------EGFHSYYTQKISELQFTVNERQKNLLRLQAQRNELNLKVRLLREE  56

Qy   74  LQLLQEPGSSYVGEVAVKVNGKNKVLVKVHPEGKYVVVDIDXNIDITKIIPSTRVALRNDSYV  133
Db   57  LQLLQEQGSSYIAEVVKPMDKNKVLVKVHPEGKYVVDVDKTINIKDVTPSSRVALRNESYT  116

Qy  134  LHIVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQ  193
Db  117  LHKILPNKVDPLVSLMLVEKVPDSTYEMVGGLRKDIQIQEIKEVIELPVKHPELFDALGITQ  176

Qy  194  PKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKYIGEGSRMVRELFVMAREHA  253
Db  177  PKGVLLYGPPGTGKTLLARAVAHHTECTFIRVSGSSELVQKFIGEGSRMVRELFVMAREHA  236

Qy  254  PSIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDIL  313
Db  237  PSIIFMDEIDSIGSARLETGT--GDSEVQRTMLELLNQLDGFBATKNIKVIMATNRIDVL  294

Qy  314  DQALLRPGRIDRKIEFPPPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKA  373
Db  295  DQALLRPGRIDRKIEFPPPNEEARLDILKIHSRKMNLTRGINLRKIAEEMPGASGAEVKG  354

Qy  374  VCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK  418
Db  355  VCTEAGMYALRERRVHVTQEDFEMAVSKVMMKDSEKNMSIRKFWK  399
```

RESULT 11

ID   Q8I4U5   PRELIMINARY;   PRT;   435 AA.
AC   Q8I4U5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Tat-binding protein homolog.
GN   PF12345C.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014851; AAN36553.1; -.
SQ   SEQUENCE   435 AA; 49545 MW; 85237FF035B94039 CRC64;

Query Match   75.1%; Score 1588.5; DB 5; Length 435;
Best Local Similarity   72.8%; Pred. No. 5.4e-95;
Matches 313; Conservative 49; Mismatches 49; Indels 19; Gaps 3;

Qy   6 VELKHAAEGVP--------------------------EANCSAKPTKQGE-GLRHYYSLNIHEHQLLLRQ 48
            :: |||| :| || || || |||||||||||||:|| ||||||
Db   8 VRSKNLSEGIKNEISEKNMENNNNNNDNNNKNSSALNEEQVQSGIKRYYELKIEEYESIINK 67

Qy   49 KTHNLNRLEAQRNDLNSRVRMLREELCLLQEPGSYVGEVKVKMGKNKVLVKVHPEGKYVV 108
          ||:|||||||||||||||||||||||:|| ||:||||||||||||||||||||||||||||
Db   68 KLQNKRLEAQRNELNARVRELCDEIQYLLEAASYVGEIVKPMGKNKVLVKINPEGKYVV 127

Qy   109 DIDKNIDIKTIKTPSTRVALRNDSYVLHLVLPSKVDPLLVNLMKVEKVPDSTYDMIGGLDQQ 168
          ||||||||||| |||||||||||||||||::|||||||||:|||||||||||||:|||||||
Db   128 DIARHINISHCPTNTRVALYNDSYKLHKILPSKVDPLVSLMKVEKVPDSTYEMVGGLDQQ 187

Qy   169 IKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGS 228
          |||||||||||:||||||||||||||||||||||||||||||||||||||:|||||||||
Db   188 VKEVKEVIELPVKHPEIFESLGISQPKGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGS 247

Qy   229 ELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARMESSGSGNGDSEVQRTMLEL 288
          |||||||||||||||||||||||||||||||||::: ||||| ||:|||||||||||:||
Db   248 ELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSQRIE--GEHGDSEVQRTMMEL 305

Qy   289 LNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEPPTPNEBSRLDILKIHSRRM 348
          |||||||||||||||||||||||||||||||||||||||||:||||:|||||||||||||
Db   306 LNQLDGFESTQNIKVIMCTNRIDILDEALLRPGRIDRKIEFPNPNVEARMELLKIHSRKM 365

Qy   349 NLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHVTQEDFEMAVAKVMKKETE 408
          ||||||||||||||||||||||||||||||||||||||||||||||:||||||||::|:
Db   366 NLMRGIDMLKIATDMNNCSGAEVKAVCTEAGMFALREERVHVTQEDFEMAVAKVMKQDAE 425

Qy   409 KNMSLRKLWK 418
          ||||||||||
Db   426 KNFTLRKLWK 435

RESULT 12

ID   Q9XTT9   PRELIMINARY;   PRT;   416 AA.
AC   Q9XTT9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Y49E10.1 protein.
GN   Y49E10.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Barlow K.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z98866; CAB11558.1; -.
DR   WormPep; Y49E10.1; CE22219.
DR   InterPro; IPR005937; 26S_p45.
DR   InterPro; IPR005939; AAA_ATPase.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR003960; AAA_sub.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01242; 26Sp45; 1.
DR   PROSITE; TIGR01242; 26Sp45; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   ATP-binding.
SQ   SEQUENCE   416 AA; 46249 MW; 8F6B840EF99B1298 CRC64;

Query Match   74.3%; Score 1572; DB 5; Length 416;
Best Local Similarity   75.7%; Pred. No. 6e-94;
Matches 306; Conservative 49; Mismatches 41; Indels 8; Gaps 2;

Qy   22 AKPTKQ-------GEGLRHYYSLNIHEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLREEL 74
            ||||| ||||||||||||:|| |||||||||||||||||||||||||||||||
Db   14 SKPTAQRLTEESDEKTLRKYFSTKVDDAQQKVADKSQNVRLQAQRNELNTKVRMLKEEL 73

Qy   75 QLLQEPGSYVGEVKVKMGKNKVLVKVHPEGKVVVDIDKNIDITKIITPSTRVALRNDSVL 134
          |||||||||||| || ||||||||||||||| |||||||:||||||||||||||||||||
Db   74 QQLHEQGSYVGEVSKAMDKKKVLVKVHPEGKIVVVDVDKSIDINSLNTGARVALRADSYAL 133

Qy   135 HLVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQP 194
          || ||||||||:|||||||||||||||:|||||||||||||||||||||:|||||||| |
Db   134 HKLLPNKVDPLVSLMKVEKVPDSTYEMVGGLDKQIKEIKEVIELPVKHPELFDALGIAQP 193

Qy   195 KGVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSELVQKYIGEGSRMVRELFVMAREHAP 254
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   194 KGVLLFGPPGTGKTLLARAVAHHTECTFIRVSGSELVQKFIGEGARMVRELFVMAREHAP 253

Qy   255 SIIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILD 314
          ||||||||||||| :|: |:|||||||||||||||||||||::|||||||||||||||||
Db   254 SIIFMDEIDSIGSSRVE-GSSGGDSEVQRTMLELLNQLDGFEATKNIKVIMATNRIDILD 312

Qy   315 QALLRPGRIDRKIEFPTPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAV 374
          |||||||||||||||| ||::||||||||:||||||||||::||||:||:||||||||:
Db   313 PALLRPGRIDRKIEFPAPDEKARADILKIHSRKMNLMRGINMAKIAEQIPGASGAEVSSV 372

Qy   375 CTEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
          |||||||||||||||||||||||||||||||:| :| ||||||
Db   373 CTEAGMFALRERRIHVTQEDFEMAVGRKVMQKDSEKNMSIKKLWK 416

RESULT 13

ID   Q9P665   PRELIMINARY;   PRT;   389 AA.
AC   Q9P665;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Probable 26S protease subunit (SUG1) protein.

GN B1D4.170.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL35928; CAB91305.1; -.
DR InterPro; IPR005928; 26S_p45.
DR InterPro; IPR005937; 26S_p45.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR003959; AAA_ATPase.
DR InterPro; IPR003960; AAA_ATPase_centr.
DR InterPro; IPR003960; AAA_sub.
DR Pfam; PF00004; AAA; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR01242; 26Sp45; 1.
DR PROSITE; PS00674; AAA; 1.
KW ATP-binding; Protease.
SQ SEQUENCE 389 AA; 43584 MW; CE13871A079CA793 CRC64;

Query Match   72.2%; Score 1528.5; DB 3; Length 389;
Best Local Similarity   78.6%; Pred. No. 3.6e-91;
Matches 305; Conservative 30; Mismatches 52; Indels   1; Gaps   1;

Qy   31 LRHYYSLNIHEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLREELQLLQEPGSYVGEVVKV 90
              |||||||||||||||||||||| |||||||||||||||||||||| ||||| ||:|||||
Db   3 LDNYYHKIEAMKLELKGQAVLRRLEAQRNDYNSRVRLLREELGLLLQQPGSSYVGEVVKV 62

Qy   91 MGKNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLHLVLPSKVDPLVNLMK 150
              |||||||||||||||||||| ||||||||||||||||||||||||| || || ||| ||
Db   63 MSTKKVLVKVHPEGKYVVDVSDNVDITKLTVGKRVTLLSDSYKLEKMLPSSVDPLVSLMM 122

Qy   151 VEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPKGVLLYGPPGTGKTLL 210
              ||| :|||||||||||||||||||||:|||||||||||||||||||||||||||||||||
Db   123 VEKVPDSTYDMIGGLDQQIKEIKEVIELGLKHPELFESLGIAQPKGVLLYGPPGTGKTLL 182

Qy   211 ARAVAAHHTDCTFIRVSGSSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSARM 270
              |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||:
Db   183 ARAVAAHHTDCKFIRVSGSSELVQKYIGEGSRMVRELFVMAREHAPSIIFMDEIDSIGSSRV 242

Qy   271 ESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQALLRPGRIDRKIEFP 330
              |||||||||||||||||||||||||||||||||||||||||||| :||||||||||||||
Db   243 E-GSSGGDSEVQRTMLELLNQLDGFEPTKNIKVIMATNRLDILDPALLRPGRIDRKIEFP 301

Qy   331 TPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVCTEAGMFALRERRVHV 390
              ::|:| ||||||||:||||||||||||||:|||||||||:|| |||||||||||||||||
Db   302 PPSVEARADILRIHSRKMNLTRGINLTKIAEKMNGCSGAELKGVCTEAGMYALRERRVHV 361

Qy   391 TQEDFEMAVAKVMKKETEKNMSLRKLWK 418
              ||||||||||| :| || | ||:||||
Db   362 TQEDFELATAKILNKHDDKEVSLAKLWR 389

RESULT 14
ID   Q9TZ65;
AC   Q9TZ65;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical 45.7 kDa protein.
GN F56F11.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Latreille P., Kramer J., Keppler D.;
RT "Direct Submission.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RT "The sequence of C. elegans cosmid F56F11.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF099222; AAK21407.2; -.
DR WormPep; F56F11.4a; CE28467.
DR InterPro; IPR005937; 26S_p45.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR003959; AAA_ATPase.
DR InterPro; IPR003960; AAA_ATPase_centr.
DR InterPro; IPR003960; AAA_sub.
DR Pfam; PF00004; AAA; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR01242; 26Sp45; 1.
DR PROSITE; PS00674; AAA; 1.
KW Hypothetical protein; ATP-binding.
SQ SEQUENCE 411 AA; 45690 MW; 2F3F7566E17926C3 CRC64;

Query Match   70.9%; Score 1501; DB 5; Length 411;
Best Local Similarity   72.5%; Pred. No. 2.3e-89;
Matches 292; Conservative 53; Mismatches 52; Indels   6; Gaps   2;

Qy   21 SAKP-----TKQGEGLRHYYSLNIHEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLREELQ 75
              |:|: |||:|||||||||||||||||||||||||||||||||||:|||||||||||
Db   10 SRKPATLDANSDEQTLRPYFKTKVEQAQEQTVAEKSINVRRLMAQQKELNGKVLNSRVRLLEI 69

Qy   76 LLQEPGSYVGEVVKVKVWKGSKNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLH 135
              ||||||||||||::: || | |||||||||||||||||||||||||||||||||||||||:|
Db   70 HLHEQQSHVGEVCKAMDKKKVLVKVHPDGKYIVDVAASVDISTLAAGTRVALRADRNYAIH 129

Qy   136 LVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVIELPIKHPELFESLGIAQPK 195
              |:||||||||||:|||||||||||||||||||||||||||||||||||||||||||||:|
Db   130 EVLPSKVDPLVSLMMVEKVPDSTYEMVGGLDTQIKEIKEVIELPLFDALLRPGRIDRKIEFP 189

Qy   196 GVLLYGPPGTGKTLLARAVAAHHTDCTFIRVSGSSELVQKYIGEGSRMVRELFVMAREHAPS 255
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   190 GVLLYGPPGTGKTLLARAVAAHHTECTFIRVSGSSELVQKFIGEGSARMVRELFVMAREHAPS 249

Qy   256 IIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQ 315
              ||||||||||||:| ||| |||||||||||||||||||||||||||||||||||||||
Db   250 IIFMDEIDSIGSSRVE-GSRGGDSEVQRTMLELLNQLDGFEATKNIKVIMATNRLDILDS 308

Qy   316 ALLRPGRIDRKIEFPTPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVC 375
              |||||||||||||| || | ||||||||:||||||||||||||||||||:||| ||||
Db   309 ALLRPGRIDRKIEFPAPDEKARAQILKIHSRKMNLMRGIRMDKIAEQIPGASGAEVKSVC 368

Qy   376 TEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
              |||||||||||:|||||||||||||||||||||||||||||||
Db   369 TEAGMFALRERRIHVTQEDFEMAVGKVMQKDSEKNMSIRKLWK 411

RESULT 15
ID   Q8MXF1;
AC   Q8MXF1;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein F56F11.4b.

```
GN   F56F11.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   Waterston R.;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
EMBL; AF099922; AAM48537.1; -.
WormPep; F56F11.4b; CE31008.
InterPro; IPR005937; 26S_p45.
InterPro; IPR003593; AAA_ATPase.
InterPro; IPR003959; AAA_ATPase_centr.
InterPro; IPR003960; AAA_sub.
Pfam; PF00004; AAA; 1.
SMART; SM00382; AAA; 1.
TIGRFAMs; TIGR01242; 26Sp45; 1.
PROSITE; PS00674; AAA; 1.
KW   Hypothetical protein; ATP-binding.
SQ   SEQUENCE   432 AA;  48003 MW;  3EB596F367FDBBC1 CRC64;
```

(Note: several RC/RN/RP/RL/RT/RA lines are partially legible in the rotated image)

```
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
RA   Latreille P., Kramer J., Keppler D.;
RT   "The sequence of C. elegans cosmid F56F11.";
```

```
Query Match         70.9%;   Score 1501;  DB 5;   Length 432;
Best Local Similarity   72.5%;   Pred. No. 2.5e-89;
Matches  292;  Conservative  53;  Mismatches  52;  Indels   6;  Gaps   2;

Qy     21 SAKP-----TKQGEGLRHYYSLNIHEHQLLLRQKTHNLNRLEAQRNDLNSRVRMLREELQ 75
                   :  |||:        :  :  :|  |||||  :|:::|:::|::: |:|::
Db     31 SRKPATLDANSDEQTLRPYFKTKVEQAEQTVAEKSINVRRLMAQQKELNGKVRLLKEEIS 90

Qy     76 LLQEPGSYVGEVKVWMGKNKVLVKVHPEGKYVVDIDKNIDITKITPSTRVALRNDSYVLH 135
             | :||:  |    | |||||  ||:|:|:       :|||:| :|:  ||:
Db     91 HLHEQGSHVGEVCKAMDKKVLVKNHPDGKYIVDVAASVDISTLAAGTRVALRADNYAIH 150

Qy    136 LVLPSKVDPLVNLMKVEKVPDSTYDMIGGLDQQIKEIKEVTELPIKHPELFESLGIAQPK 195
             :|||||:|| :|  |||||:|:  :|||||  |||||||||||:|:||||:|||||||
Db    151 EVLPSKVDPLVSLMWVEKVPDSTYEMVGGLDTQIKEIKEVTELPVKHPELFDALGIAQPK 210

Qy    196 GVLLYGPPGTGKTLLARAVAHHTDCTFIRVSGSSELVQKYIGEGSRMVRELFVMAREHAPS 255
          ||||||||||||||||||||||||:|||||||:||||||||||||||||||||||||||:
Db    211 GVLLYGPPGTGKTLLARAVAHHTECTFIRVSGSELVQKFIGEGARMVRELFVMAREHAPS 270

Qy    256 IIFMDEIDSIGSARMESGSGNGDSEVQRTMLELLNQLDGFEASNKIKVLMATNRIDILDQ 315
          ||||||||||||::|::| :  :||:||:|||||||||||||:: || |:|||||||||
Db    271 IIFMDEIDSIGSSRVE-GRGGDSEVQRTMLELLNQLDGFEATKNIKVIMATNRIDILDS 329

Qy    316 ALLRPGRIDRKIEFPTPNEESRLDILKIHSRRMNLMRGIDLKKIAEKMNGASGAELKAVC 375
          |||||||||||||:|  |:||| |||| |||:|| ||||:|| || ||:|||:| |:||
Db    330 ALLRPGRIDRKIEFPAPDEKARAQILKIHSRKWNLMRGIRMDKIABQIPGASGAEVKSVC 389

Qy    376 TEAGMFALRERRVHVTQEDFEMAVAKVMKKETEKNMSLRKLWK 418
          |||||||||||:||||||||||:|||||: :::|||||||||||
Db    390 TEAGMFALRERRIHVTQEDFEMAVGKVWQKDSEKNMSIKKLWK 432
```